



AI and Protecting Data

Ulf Mattsson
Chief Security Strategist
www.Protegrity.com

Topics:

Advances in AI

Different practical data protection models, including pseudonymization, anonymization, tokenization, encryption, and more

GDPR and recent approaches for Data Transfer

How to reduce the Risk of Ransomware



A I and m ore



Distinctions

- The term “AI” has been used to describe many things.
 - NLP (Alexa, Google)
 - Linear/Polynomial Regression Analysis
 - Probability
 - Neural Networks (Remember the Terminator)
 - Machine Learning
 - Stochastic based models (LLM including ChatGPT)

What about Chat GPT and other LLM?

- Chat GPT is a “Large Language Model” (LLM) or Foundational Machine Learning Model. Original models built on Stochastics (some call the stochastic parrots).
- Conversational chatbot with Generative Pretrained Transformer (GPT).
- Is “generative” AI meaning, it crafts a response with “new” content and tries to format it into natural language.
- **(Stochastics** is the study of data sets with random probability distributions that can be analyzed statistically but not predicted.) Due to the uncertainty present in a stochastic model, **the results provide an estimate of the probability of various outcomes.**
- Interesting, entertaining and wrapped in a lot of hype (anyone remember the metaverse).
- GARTNER: Is ChatGPT artificial general intelligence? No.

How does it work?



- Classifies “intent” like Alexa or Google with confidence scores (statistics) BUT using stochastic-like analysis.
- Produces “constraints” to bound the response.
- Trained with up to 300 Billion Words from various sources.
- Can summarize responses with marginal degrees of accuracy (use cautiously) conditional upon input.
- Models are fine-tuned by your feedback (unless you use the API)
- Generates outputs based on trained foundational models (i.e. If the model is not trained in a particular area, it doesn't work).
- Uses probability analysis.
- Determine the best (most probable) path based on your input.

More specifically, model strengths include



- Generate and augment prose or narratives
- Code development, translation, explanation and augmentation
- Summarize and simplify long-form texts.
- Classify content for sentiment or by topic area.
- Answer questions,
- Translate and convert language (including programming languages).
- Written content augmentation and creation.

What it isn't

- Accurate much of the time.
- Equally strong across all domains...only where it's trained.
- Sentient (Perceptive)...it is not AI
- Insightful i.e.) **Gives you the same answer if you ask how to build a high-performance team of plumbers or brain-surgeons.**
- Reliable & Trustworthy (ie Requires expert review).
- Able to be customized or trained with your data.
- Not particularly insightful much of the time. Regurgitates prescribed paths through the model.

Popular Use Cases of ChatGPT

ChatGPT Capabilities

- ✓ Create written content.
- ✓ Answer questions (noncomputational) and discover information.
- ✓ Transform the tone, formality or writing genre of language on request.
- ✓ Summarize and classify text.
- ✓ Compare paragraphs and correct grammar.
- ✓ Generate ideas, suggestions and key points on different topics.
- ✓ Classify and categorize content based on the example provided.
- ✓ Generate, translate, explain and verify computer code.
- ✓ Translate text to instructions, query or different language.

Select Enterprise Use Cases of ChatGPT



What Exactly is GenAI in a Professional Context?

GenAI Creates & Learns

Gartner's AI Definition:

- **Analyzes** data with logic-based techniques like Machine learning (ML)
- **Interprets** events, supports and automate decisions (careful here).

Gartner's Generative AI Definition:

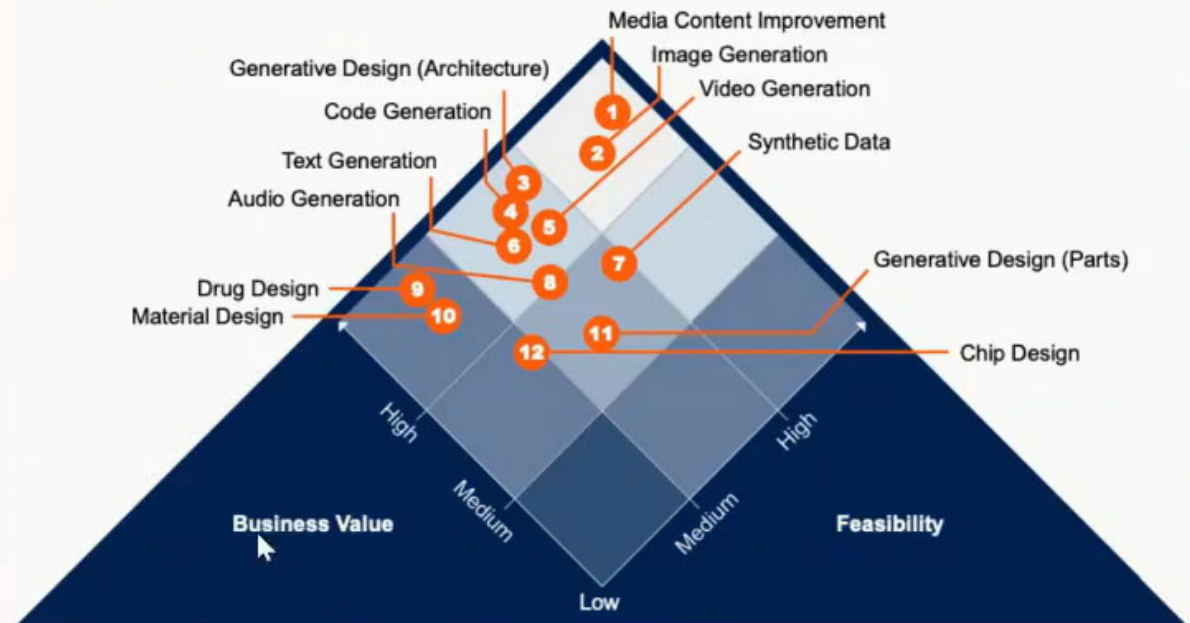
- **Creates** newly derived content, strategies, designs and methods.
- **Learns** from large repositories of original source content.

Risks Executives Should be Watching

- ⚠ Hallucinations
- ⚠ No Attribution
- ⚠ Data Leakage

What Use Cases Are Emerging for CXOs?

Gartner Use Case Prism for Generative AI



But...

- Too many organizations are jumping into the technology without understanding the problem and the use case
- This is going to create failed POCs
- Many organizations have a solution looking for a problem.
- Driven by leadership and the hype cycle.
- Have complete misunderstanding of how the models are built, use cases and limitation.

Getting Started in Gen AI Pilot

- What is your use case?
- Will OOTB model suffice?
 - Prompt Engineering & Token Filtering?
 - Model Selection?
 - Text Based UI (Chat GPT)
 - API's and Application Embedding.
- Is Model Augmentation required?
 - Model Selection?
 - Model Training, Testing and Feedback?
 - APIs used for Training?

Key Findings

- The most successful pilots focus on demonstrating business potential, not on technical feasibility. Organizations tend to run technical pilots that simply demonstrate that it is possible to build something with generative AI, leading to only incremental improvements and ignoring the transformative potential of this technology.
- IT leaders struggle to identify and prioritize impactful generative AI use cases due to the broad and emerging nature of the technology.
- Mature AI organizations involve business partners and software engineers as key members of their AI projects and pilot teams.

Enterprise ChatGPT/GPT Usage Areas: Pros and Cons

ChatGPT

Out-of-the-Box
Model Usage



Prompt
Engineering/
In Context Learning



GPT3
Model

Deployment/
Fine Tuning
of Custom
Models

Out-of-the-Box Model Usage

- This form of usage is by far the most accessible and common today.
- Text-based webchat interface (chat.openai.com). API recently available.
- For most use cases, output must be reviewed by a human, as it may contain inaccuracies or unacceptable content.
- Enterprises may achieve useful results with limited investments and skills. But because many users are inexperienced, they risk overlooking data, security and analytics risks.
- A limitation is that the model cannot include real-time, current or custom data. Nor does it cover recent historical events (those after December 2021). However, new data can be added via a prompt at the time of interaction.

Prompt Engineering/In Context Learning

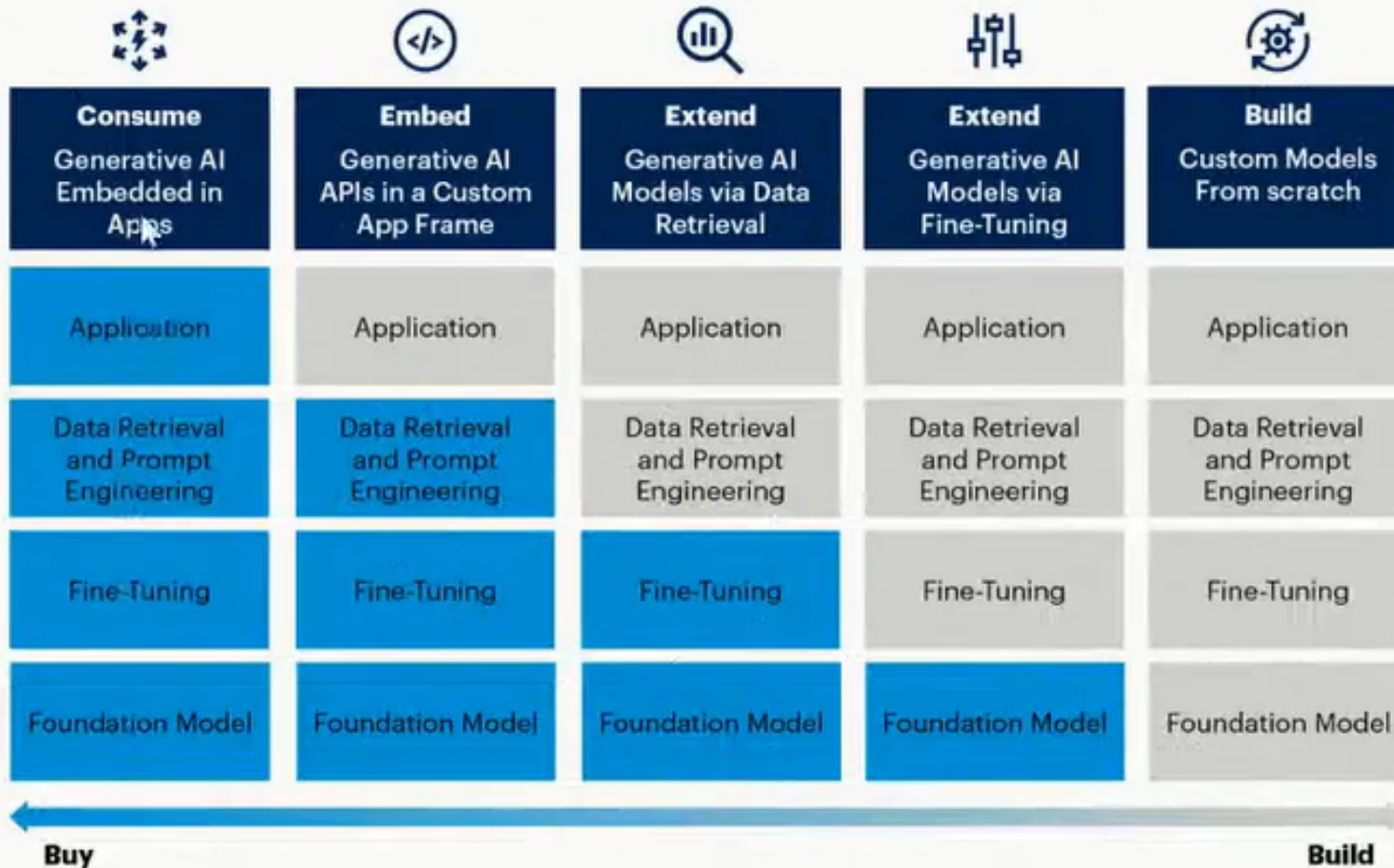
- Prompt engineering can be applied to both ChatGPT and GPT use cases. It involves developing a systematic approach to creating, tuning, and evaluating results in terms of inputs and outputs to and from ChatGPT.
- In ChatGPT, the prompt is the critical element driving results. Small changes to a prompt's choice of words and word order can result in significant changes in output. A prompt can also contain data that should be incorporated or considered when generating a response.
- Leaders should anticipate that prompt engineering is a new technical skill that will need to be developed, along with related tools.
- In some cases, this requirement will extend to building a separate learning model to optimize prompts.
- In Context Learning, leveraging Retrieval Augmented Generation, is the dominant model in use by organizations that must keep data secure and regularly update data in an LLM context

Deployment/Fine Tuning of Custom Models

- This is the likely long-term approach for sophisticated solutions.
- This approach is not possible with ChatGPT, as it does not provide users with access to customize its underlying model.
- Besides GPT, other foundation models exist. Some are specialized.
- Customizing foundation models is a complex task that requires significant skills, data curation and funding.
 - Enterprises should anticipate a robust market for third-party models customized for different use cases.
 - Planners should anticipate the emergence of third-party, fit-for-purpose, specialized models. Buying one of these may prove a better approach for many enterprises than customizing a model themselves.
 - Applications may also have prebuilt models for their users.

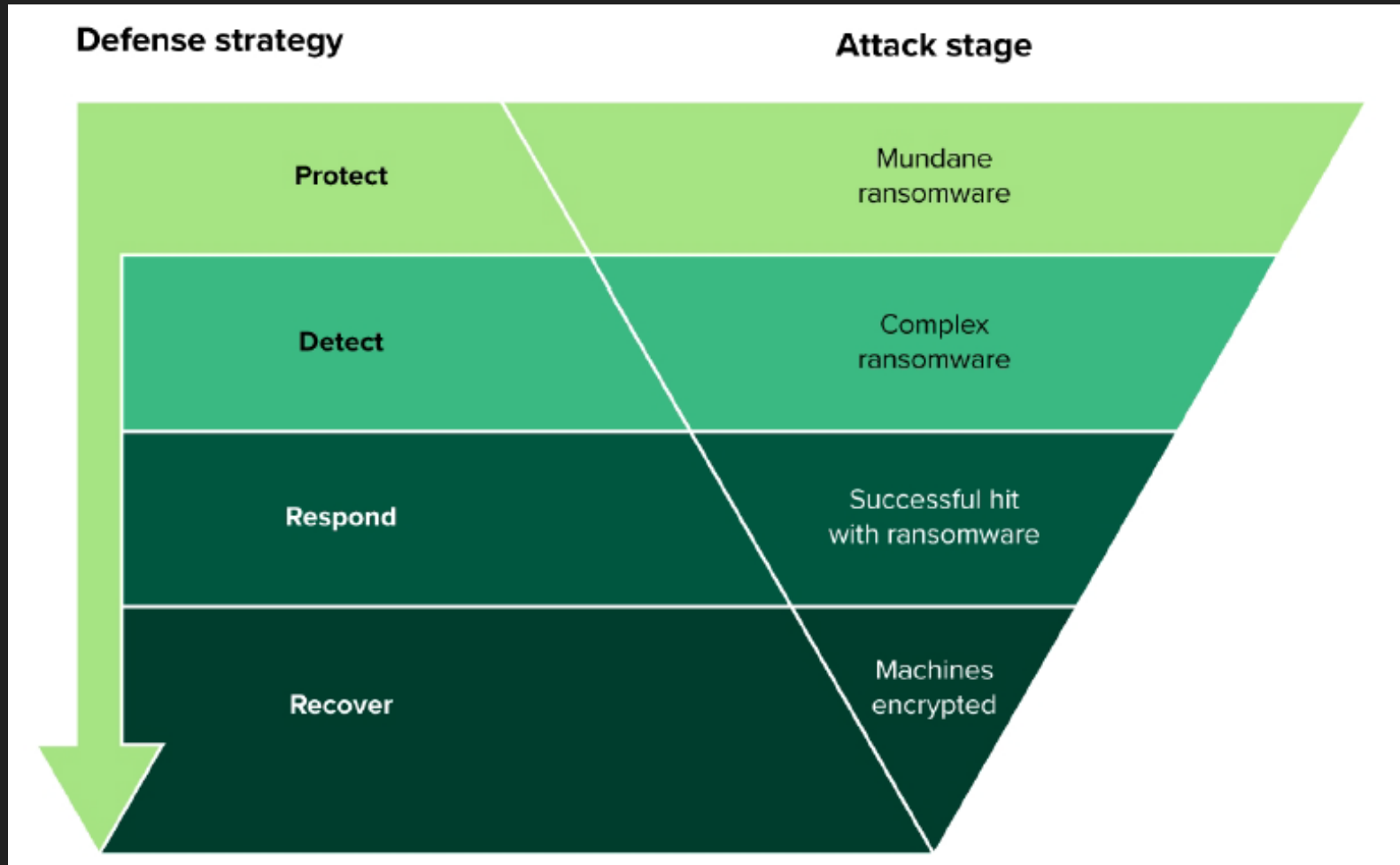
Generative AI Deployment Approaches

■ Provider-Managed ■ Self-Managed



Reducing the Risk of Ransomware

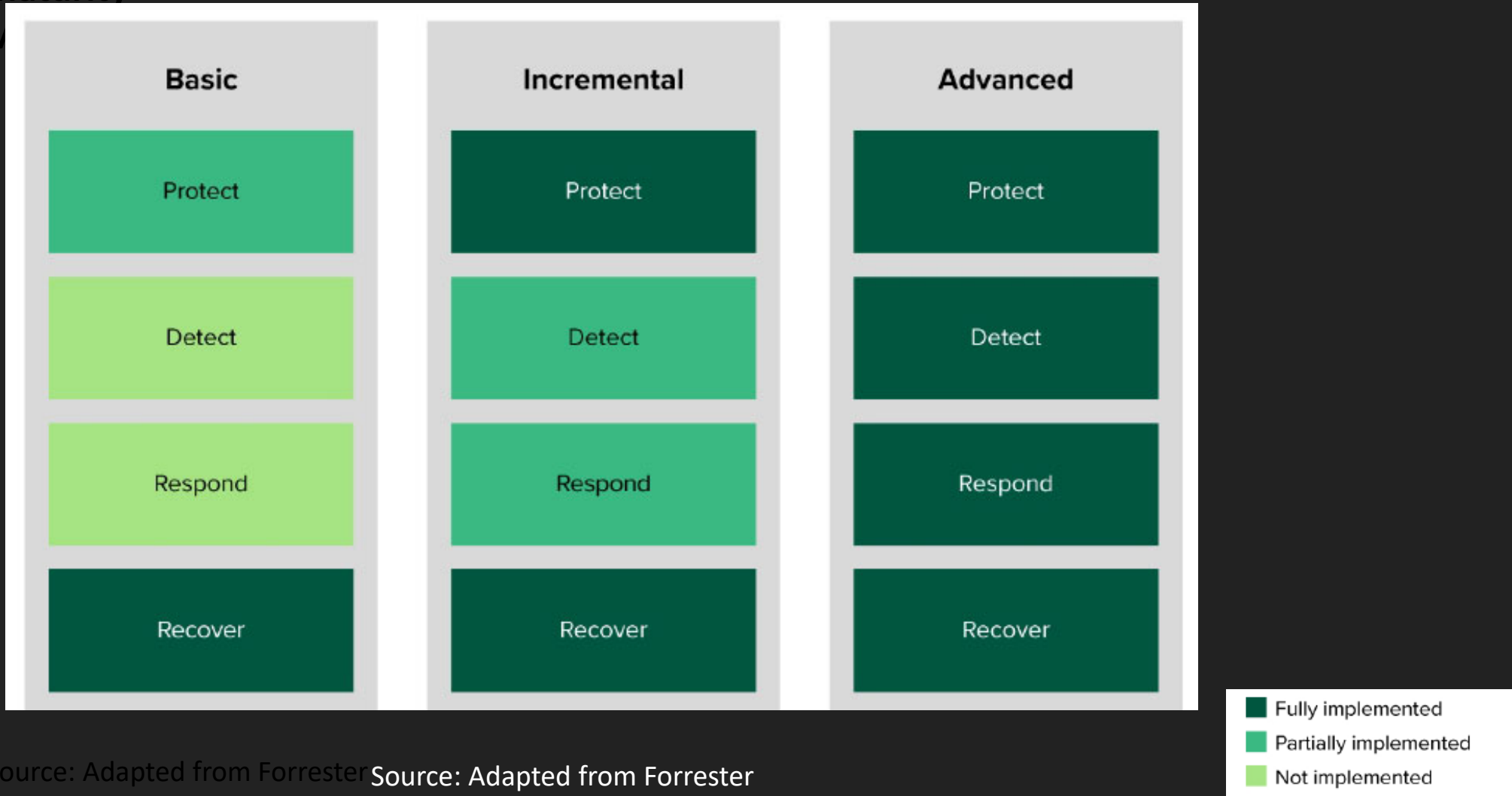
A Framework for Aligning Ransomware Defenses to Attacks



Source: Adapted from Forrester

Ransomware Maturity

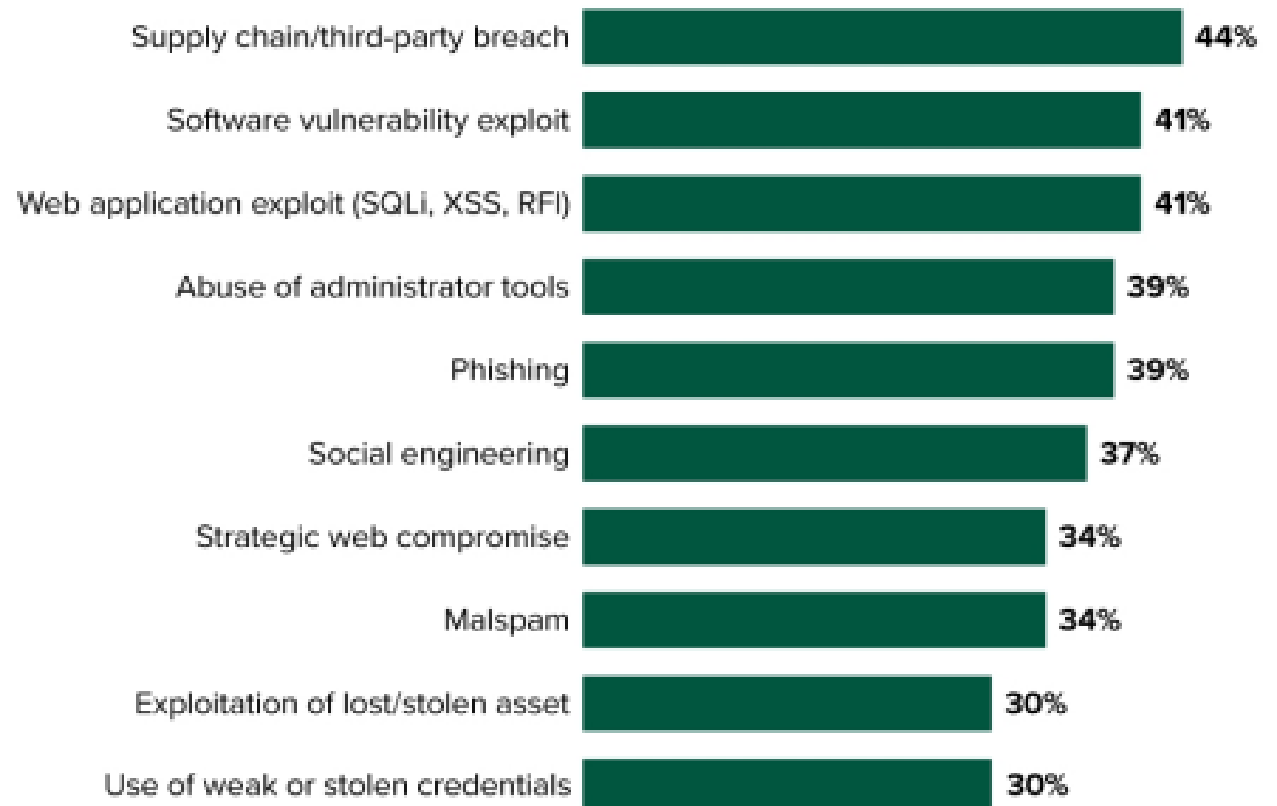
The State-Of-Ransomware-Attacks-And-Defenses



Source: Adapted from Forrester Source: Adapted from Forrester

Ransomware Modes of Attacks

"How was the external attack carried out?"



Source: Adapted from Forrester Source: Adapted from Forrester

The Reasons for investing in preparation against Ransomware attacks

Reputation



Depending on the type of service a company provides, an attack can pose an existential threat. Service failures or the leakage of customer data can damage the brand.

Quantity



There is an increasing number of individuals that specialize in cybercrime. The number of serious attacks is increasing every year and the chance of falling victim to an attack is significantly high.

Digitalization



The trend towards 100% digitization leads to a dependence on IT systems and applications.

From the healthcare to the manufacturing industry, rarely are processes still manual and in the event of a failure, no further work can be done.

Hybrid



Based on digitization, more and more different operating models are emerging from the classic on-premises to hybrid and cloud-native.

This causes interfaces that are potentially accessible via the Internet, or access to internal company data via portals.

Costs



Costs of blackmailing that are better left unpaid.

Loss of customers or decline in orders.

Costs for forensics and the commissioning of these systems.

Loss of expert knowledge from the company (construction plans, source code, etc.)

Source: Adapted from Forrester

GDPR and Data Transfer

The adequacy decision claims to offer more protection for EU data that is transferred to the US.

In a major move for data transfer capabilities, The European Commission has adopted its adequacy decision for the EU-US Data Privacy Framework.

The decision by the EU has concluded that the United States ensures an adequate level of protection – comparable to that of the European Union – for personal data transferred from the EU to US companies under a new framework.

According to the EU Commission, the framework introduces new binding safeguards to address all the concerns raised by the European Court of Justice, including limiting access to EU data by US intelligence services to what is necessary and proportionate, and establishing a Data Protection Review Court (DPRC), to which EU individuals will have access.

The EU says that this decision offers "significant improvements" compared to the Privacy Shield, which existed up until the Shrems II decision sparked by whistleblower Edward Snowden which revealed the surveillance of people's digital data by US authorities.

The revelation triggered Max Shrems, a data privacy advocate, to call on the EU to withdraw the adequacy agreement with the US as EU citizen data could be under surveillance by the US state, which was against the EU's GDPR.

<https://www.digit.fyi/landmark-eu-us-data-privacy-framework-adopted/>

The court will independently investigate and resolve complaints, including by adopting binding remedial measures.

The safeguards put in place by the US will also facilitate transatlantic data flows more generally, since they also apply when data is transferred by using other tools, such as standard contractual clauses and binding corporate rules.

While this is a major step towards streamlined international data transferred, the Framework will be subject to periodic reviews carried out by the European Commission, together with representatives of European data protection authorities and competent US authorities.

The first review will take place within a year of the entry into force of the adequacy decision, in order to verify that all relevant elements have been fully implemented in the US legal framework and are functioning effectively in practice.

Privacy advocates remain unconvinced by the new framework. Activists at NYOB ([None of Your Business](#)), the data privacy advocacy group, have said there is little to no difference in the new framework from past EU-US agreements including Safe Harbour and the Privacy Shield, both of which did not protect EU data from US surveillance.

The Threat Landscape

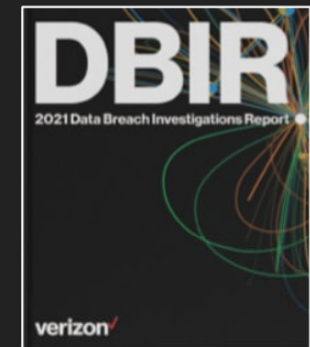
Prevent Attacks

Recover after Attacks

Defend Users and Resources

Secure Data

Secure Applications



The Threat Landscape

Phishing

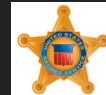
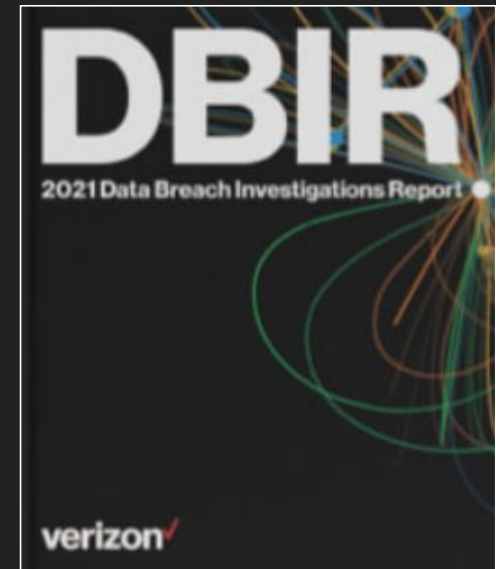
Stolen creds

Mis-delivery

Misconfiguration

Password dumper

Ransomware



United States Secret Service



FEDERAL TRADE COMMISSION

Ransom ware Pricing

The price table provided by a Chinese actor behind ransomware attacks

NUMBER OF COMPUTERS	PRICE/COMPUTER (IN USD)
1-9	3000
10-49	1500
50-99	1120
100-499	750
500-999	560
1000-4999	380
5000-9999	260

Source: NCC Group

Ransom ware Pricing

- NSS Group concluded that there are clear signs that adversaries have adopted price discrimination techniques based on the **yearly revenue of their victims.**
- If we look at the price setting and negotiation from the adversaries' point of view, we see that they wield a **massive advantage over their victims.**
- Not only do they have the luxury of investigating their **victims' financial statements** if they choose to do so, but they also have the advantage of having previous experiences they can use.

The Ransom and Revenue

Smaller companies generally pay more from a RoR* point of view.

A smaller company pays less in absolute amount but higher in percentage of their revenue.

Ransom Paid	Estimated Revenue (in Million USD)	RoR
\$14,400,000	\$17,500	822
\$1,500,000	\$1,000	1,500
\$500,000	\$1,000	500
\$350,000	\$16	21,875

Source: NCC Group

*: To calculate the RoR (Ransom per annual Revenue), we divided the ransomware demand by the annual revenue a company made in the last year before they got attacked.

Ransomware gang's script shows exactly the files they're after

As we would expect, the script seeks out files related to the companies financials or personal information, such as audit, banking information, login credentials, tax forms, student information, social security numbers, and SEC filings.

However, it also looks for more intriguing keywords that could be particularly harmful to a company if leaked, such as folders containing the words 'crime', 'investigation', 'fraud', 'bureau', 'federal', 'hidden', 'secret', 'illegal', and 'terror.'

The full list of 123 keywords targeted by the threat actors' script is listed in the table below.

941	confident	Info	RRHH
1040	Crime	insider	saving
1099	claim	Insurance	scans
8822	Terror	investigation	sec
9465	Confidential*Disclosure	IRS	secret
401K	contact	ITIN	security
4506-T	contr	K-1	studen
ABRH	CPF	letter	seed
Audit	CRH	List	Signed
Addres	Transact	Login	sin
agreem	DDRH	mail	soc

<https://www.bleepingcomputer.com/news/security/ransomware-gangs-script-shows-exactly-the-files-theyre-after/#.YSZRQltqa6w.linkedin>

RANSOMWARE RISK MANAGEMENT

Prevent Attacks

- Maintain antivirus & patching
- Allow only authorized apps
- Block ransomware sites
- Limit personally owned devices
- Limit administrative privileges
- Limit personal apps
- Avoid unknown files or links

Recover after Attacks

- Follow recovery plan
- Isolated backups offline, test backups
- Verify emergency contacts



NIST
CYBER

- NIST IR 8374
CYBERSECURITY
FRAMEWORK FOR
RANSOMWARE

RANSOMWARE RISK MANAGEMENT

Steps that organizations can take now to help recover from a future ransomware event include:

**Recover
after
Attacks**

1. Develop and implement an incident **recovery plan** with defined roles and strategies for decision making.
2. Carefully plan, implement, and test a data backup and restoration strategy—and secure and **isolate backups** of important data.
3. Maintain an up-to-date list of internal and external **contacts**

The logo for NIST Cyber, with "NIST" in white and "CYBER" in blue, set against a black background.

- NIST IR 8374
CYBERSECURITY
FRAMEWORK FOR
RANSOMWARE

RANSOMWARE RISK MANAGEMENT

Defend Users and Resources

- Data security policies and rules to protect data at rest and in transit
- Managing enterprise users and continuously monitors security behavior
- Protect servers and other devices

The logo for NIST CYBER, featuring the text "NIST" in white and "CYBER" in blue, both in a bold, sans-serif font.

- NIST SP 800-207
Zero Trust
Architecture (ZTA)



NIST SP 800-207 Zero Trust Architecture (ZTA)

Enterprise device security characteristics:

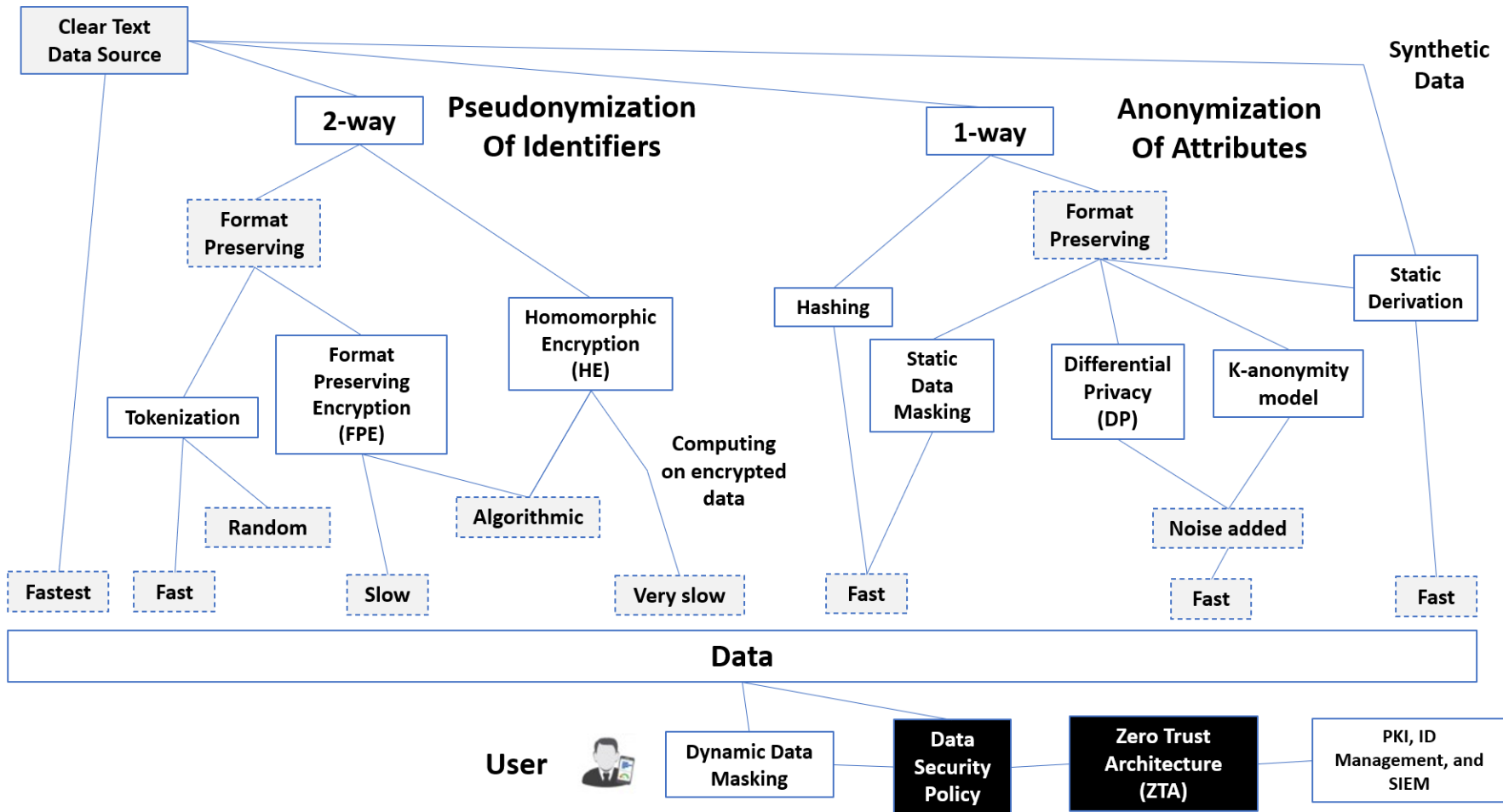
Defend Users and Resources

1. Maintaining **data protection at rest and in transit**
2. Remediating device vulnerabilities that could result in **unauthorized access to Data** stored on or accessed by the device, and misuse of the device
3. Mitigating **malware** execution on the device that could result in unauthorized access to data stored on or accessed by the device, and misuse of the device
4. Mitigating the **risk of data loss** through accidental, deliberate, or malicious deletion or obfuscation of data stored on the device
5. Maintaining awareness of and responding to suspicious or **malicious activities** within and against the device to prevent or detect a compromise of the device

<https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-207.pdf>

<https://www.nccoe.nist.gov/sites/default/files/library/project-descriptions/zta-project-description-final.pdf>

Data Protection Techniques and Zero Trust Architecture (ZTA)



RISK MANAGEMENT

Secure
Data

- Pseudonymization & Cryptographic tools
- Suppression & Generalization
- Randomization & Privacy models



- ISO / IEC 20889
INTERNATIONAL
STANDARD

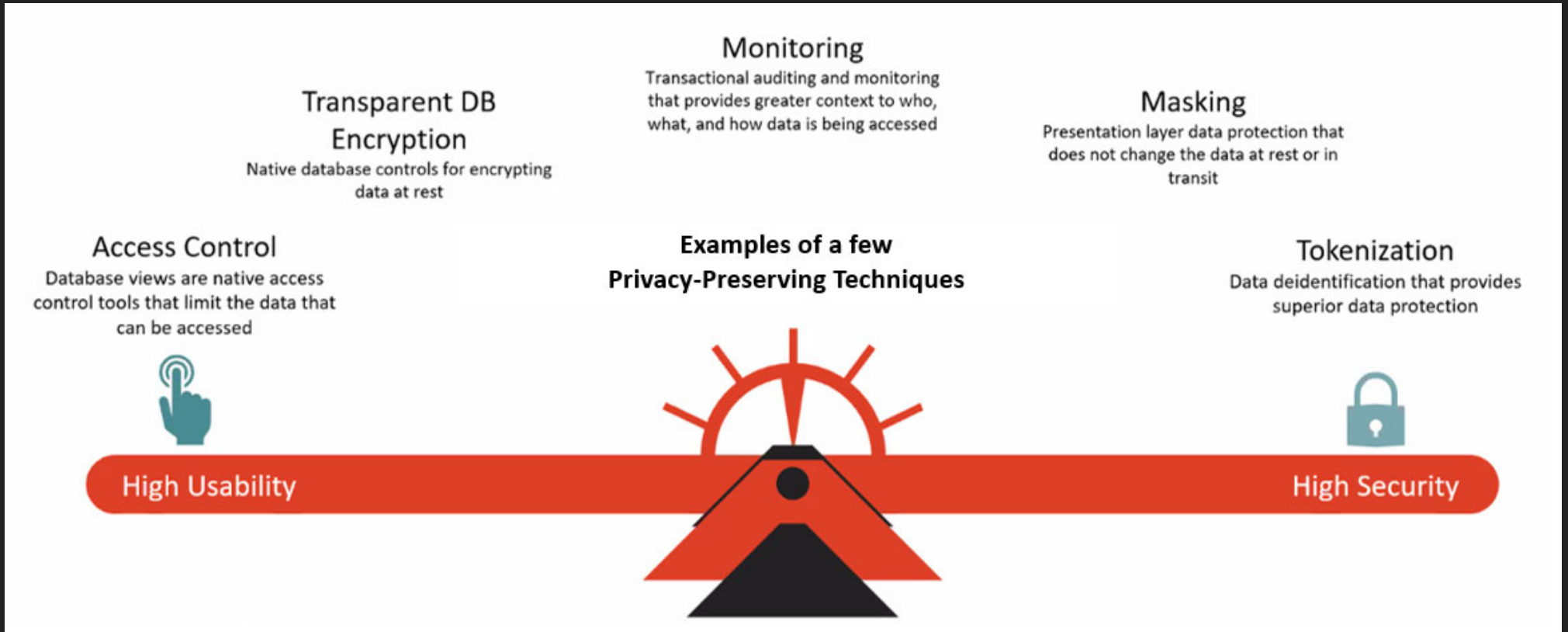


High Usability



High Security

What Data Protection Technique do I need?

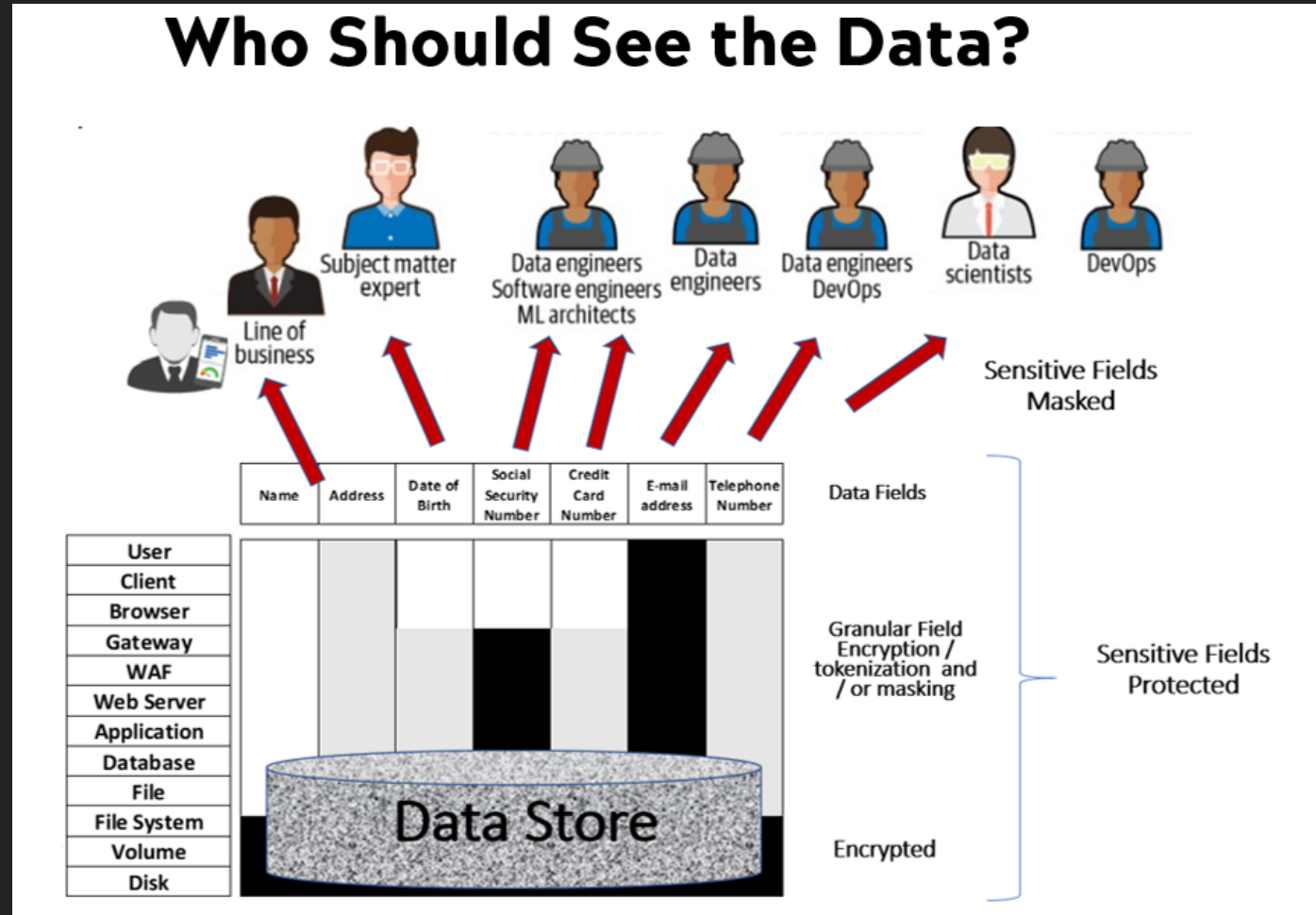


Specify Access Control and Data Protection to Use

Review Use Cases and Types of Data

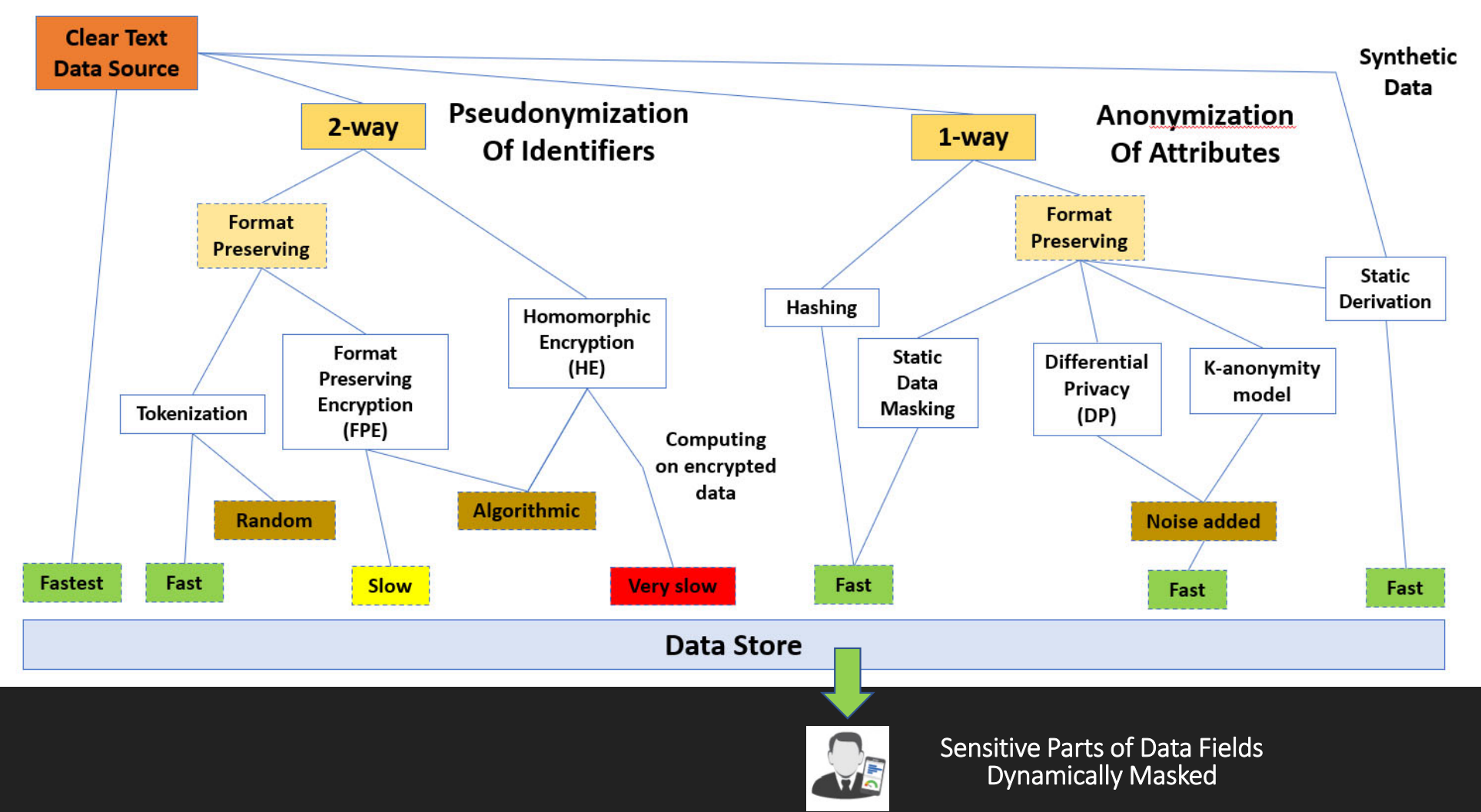
Implement

1. Dynamic Masking
2. Tokenization
3. Encryption



**Pseudonymization,
anonymization,
tokenization,
encryption, and
more**

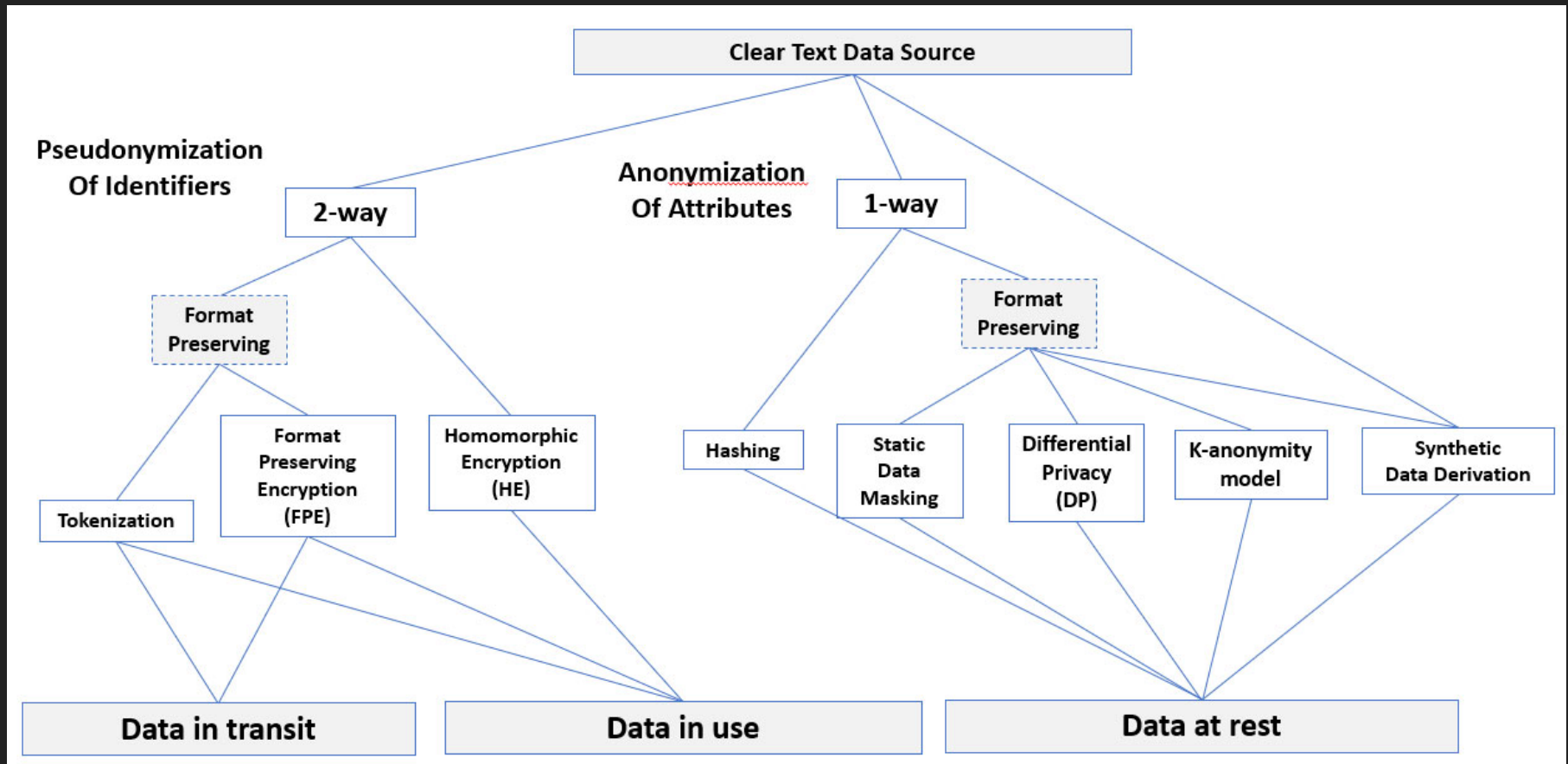
Positioning of Different Data Protection Techniques



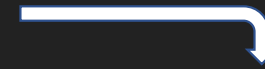
Data protection techniques: Deployment on-premises, and clouds

Privacy enhancing data de-identification terminology and classification of techniques			Data Warehouse	Centralized	Distributed	On-premises	Public Cloud	Private Cloud
De-identification techniques	Tokenization	Vault-based tokenization		y				y
		Vault-less tokenization	y	y	y	y	y	y
	Cryptographic tools	Format preserving encryption		y	y	y	y	y
		Homomorphic encryption			y		y	
	Suppression techniques	Masking	y	y	y	y	y	y
		Hashing	y	y	y	y	y	y
Formal privacy measurement models	Differential Privacy	Server model	y	y	y	y	y	y
		Local model	y	y	y	y	y	y
	K-anonymity model	L-diversity	y	y	y	y	y	y
		T-closeness	y	y	y	y	y	y

How are Data Protection Techniques different?



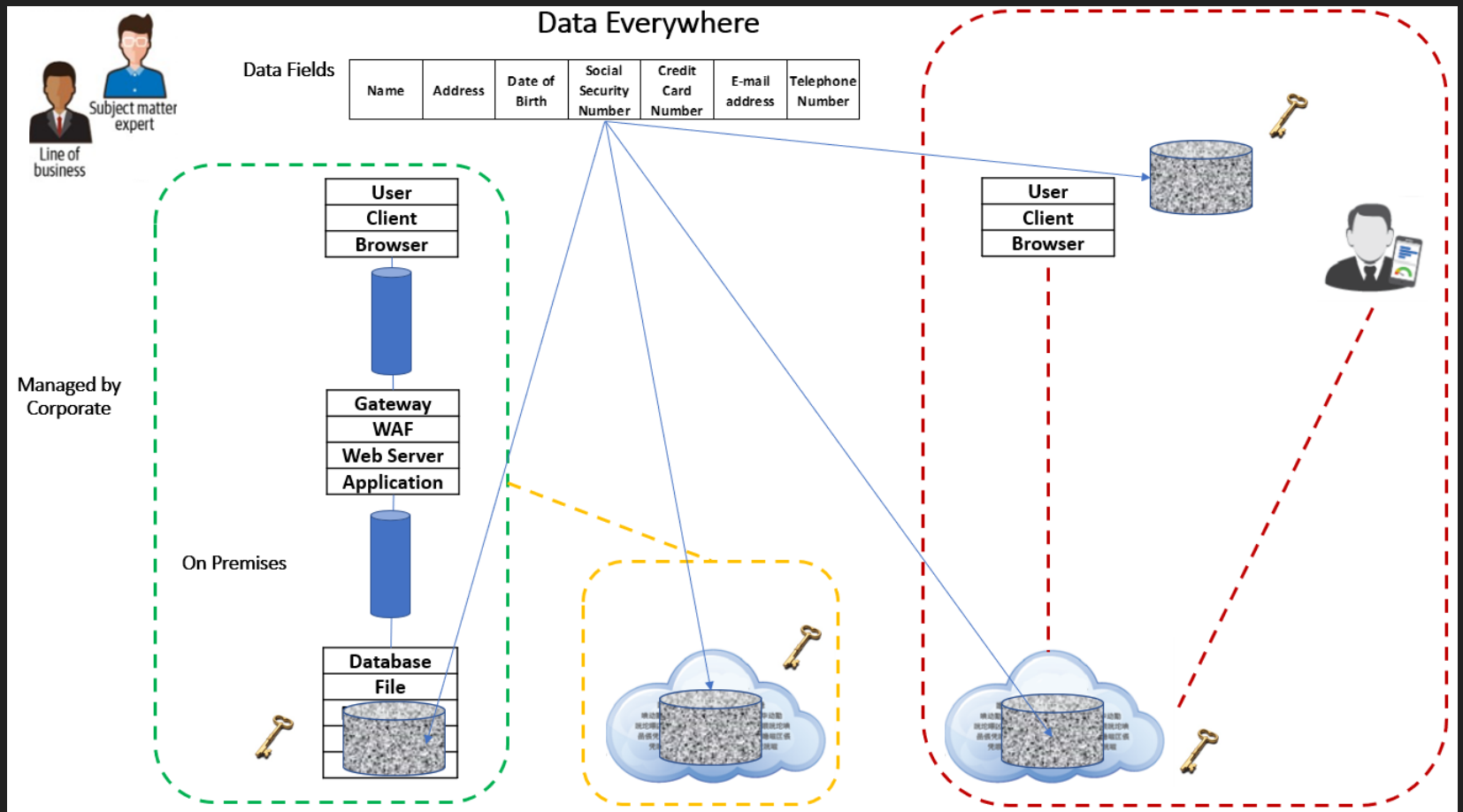
Risk Reduction



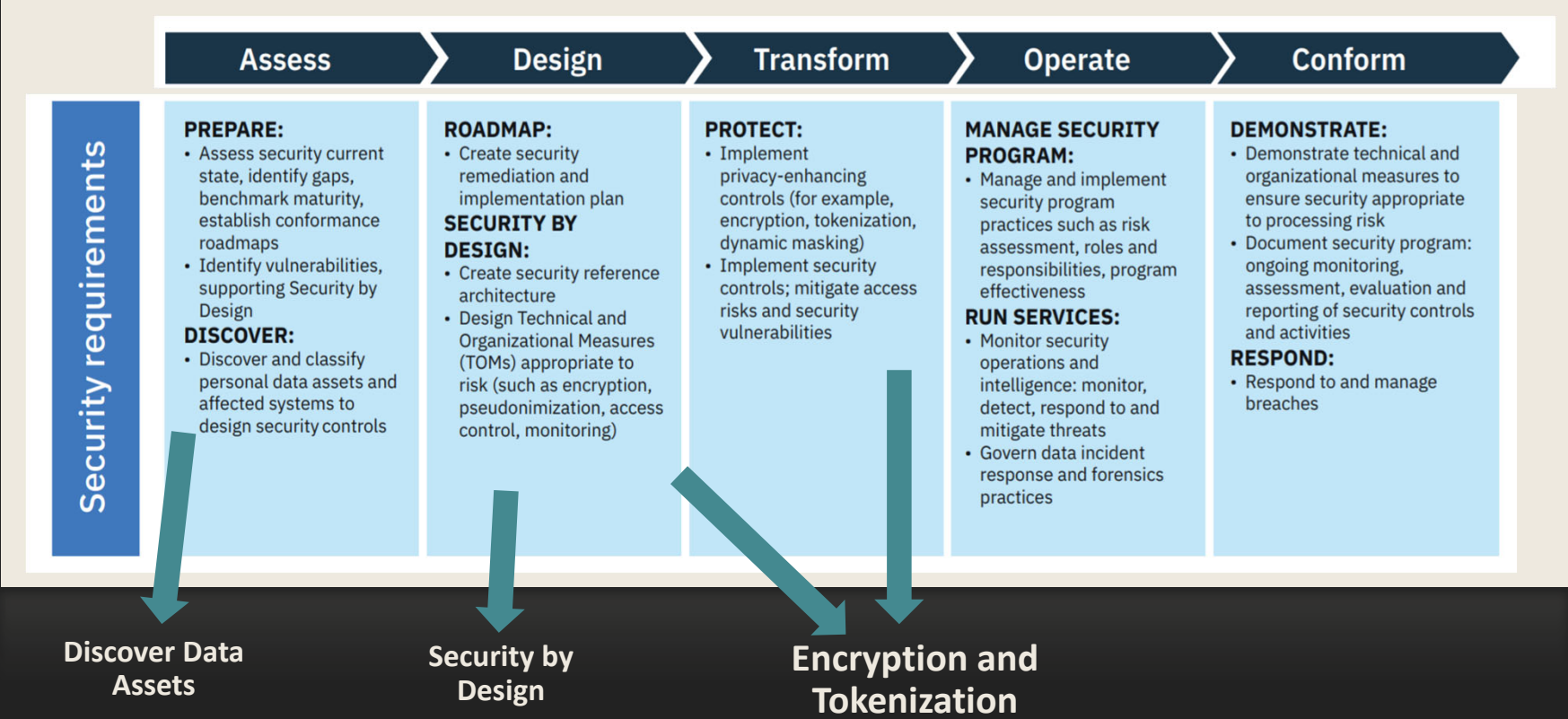
De-identification technique		Use case			Truthful at records	Applicable	Reduces risk		
		Transit	In use	At rest			Singling out	Linking	Inference
Pseudonymization	Tokenization	Yes	Yes	Yes	Yes	Direct identifiers	No	Part	No
Cryptographic tools	Deterministic encryption	Yes	No	Yes	Yes	All attributes	No	Part	No
	Order-preserving encryption	Part	Part	Part	Yes	All attributes	No	Part	No
	Homomorphic encryption	Yes	Yes	Yes	Yes	All attributes	No	No	No
Suppression	Masking	Yes	Yes	Yes	Yes	Local identifiers	Yes	Part	No
	Local suppression	Yes	Yes	Yes	Yes	Identifying attributes	Part	Part	Part
	Record suppression	Yes	Yes	Yes	Yes	All attributes	Yes	Yes	Yes
	Sampling	Part	Part	Part	Yes	All attributes	Part	Part	Part
Generalization	Generalization	Yes	Yes	Yes	Yes	Identifying attributes	Part	Part	Part
	Rounding	Yes	Yes	Yes	Yes	Identifying attributes	No	Part	Part
	Top and bottom coding	Yes	Yes	Yes	Yes	Identifying attributes	No	Part	Part
Randomization	Noise addition	Yes	Yes	Yes	No	Identifying attributes	Part	Part	Part
	Permutation	Yes	Yes	Yes	No	Identifying attributes	Part	Part	Part
	Microaggregation	Yes	Yes	Yes	No	All attributes	No	Part	Part
Privacy models	Differential privacy	No	Yes	Yes	No	Identifying attributes	Yes	Yes	Part
	k-Anonymity	No	Yes	Yes	Yes	Quasi identifiers	Yes	Part	No

Our New Distributed Environment

Where are the Encryption Keys and the Data?



GDPR Security Requirements Framework



Source: IBM

SECURITY AUDIT AND LEADERSHIP SERIES

Ulf Mattsson

Controlling Privacy and the Use of Data Assets

Who Owns the New Oil?

 CRC Press
Taylor & Francis Group

Contents

Introduction

1. Section I. Introduction and Vision

- Chapter1. Privacy, Risks, and Threats
- Chapter2. Trends and Evolution
- Chapter3. Best Practices, Roadmap, and Vision

2. Section II. Data Confidentiality and Integrity

- Chapter4. Computing on Encrypted Data
- Chapter5. Reversible Data Protection Techniques
- Chapter6. Non-Reversible Data Protection Techniques

3. Section III. Users and Authorization

- Chapter7. Access Control
- Chapter8. Zero Trust Architecture

4. Section. IV. Applications

- Chapter9. Applications, Privacy by Design, and APIs
- Chapter10. Machine Learning and Analytics
- Chapter11. Secure Multi-party Computing
- Chapter12. International Unicode Data
- Chapter13. Blockchain and Data Lineage

5. Section V. Platforms

- Chapter14. Hybrid Cloud, CASB, and SASE
- Chapter15. HSM, TPM, and Trusted Execution Environments
- Chapter16. Internet of Things
- Chapter17. Quantum Computing

Summary

Appendices

- Appendix A. Standards and Regulations
- Appendix B. Governance, Guidance, and Frameworks
- Appendix C. Discovery and Search
- Appendix D. Digital Commerce, Gamification, and A.I.
- Appendix E. Innovation and Products
- Appendix F. Glossary

Ulf.Mattsson

@

Protegrity.com

<https://www.routledge.com/Controlling-Privacy-and-the-Use-of-Data-Assets-Who-Owns-the-New-Oil/Mattsson/p/book/9781032039121>

My books:

- This is my first book at <https://www.routledge.com/Controlling-Privacy-and-the-Use-of-Data-Assets-Who-Owns-the-New-World/Mattsson/p/book/9781032039121>
- This is my second book at <https://www.taylorfrancis.com/books/mono/10.1201/9781003254928/controlling-privacy-use-data-assets-volume-2-ulf-mattsson>

