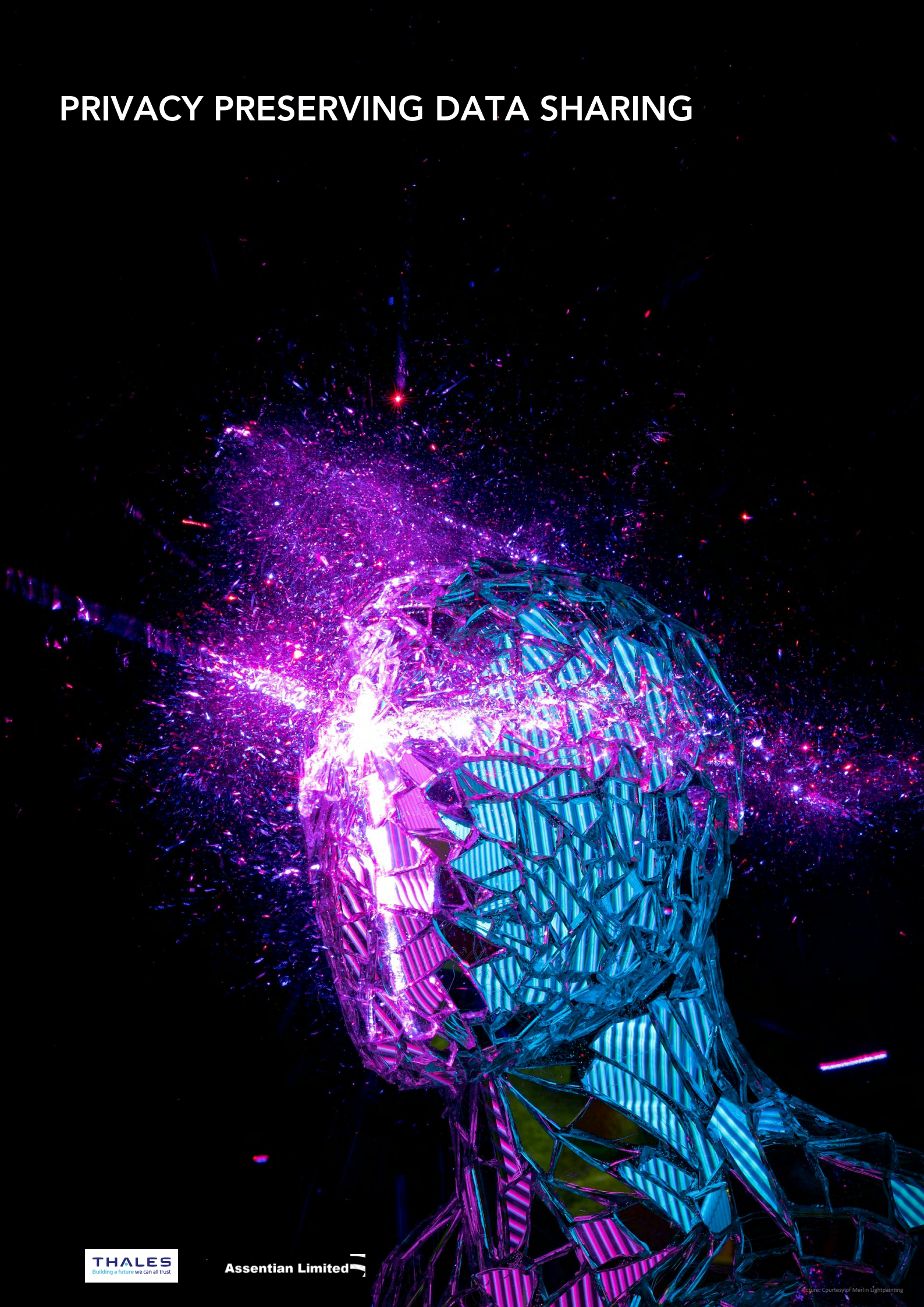
PRIVACY PRESERVING DATA SHARING



THALES
Building a future we can all trust

Assentian Limited

Picture: Courtesy of Merlin Lightpainting

A NEW ERA OF DATA SHARING WITHOUT THE RISKS INHERENT IN SHARING DATA

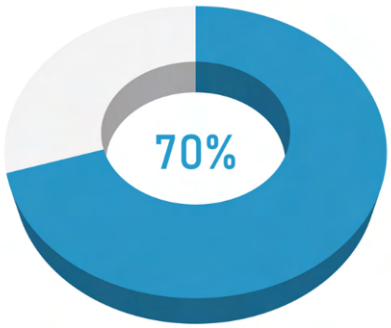
Cutting edge technologies will unlock the, as yet, unrealised value of data. Providing safeguards for AI and quantum computing, for the good of humankind, like never before

Contents

Background

Data is a vital resource for solving society's biggest problems. Today, significant amounts of data are accumulated everyday fuelled by widespread data generation methods, new data collection technologies, faster means of communication, and more accessible cloud storage. Advances in computing have significantly reduced the cost of data analytics and artificial intelligence, making it even easier to use this data to derive valuable insights and enable new possibilities. However, this potential is often limited by legal, policy, technical, socioeconomic, and ethical challenges involved in sharing and analysing sensitive information. These opportunities can only be fully realized if strong safeguards that protect privacy are in place which is seen as a fundamental right in democratic societies.

Privacy-preserving data sharing and analytics (PPDSA) methods and technologies can unlock the beneficial power of data analysis while protecting privacy. Using privacy enhancing technologies (PETs), PPDSA solutions include methodological, technical, and sociotechnical approaches that employ privacy-enhancing technologies to derive value from, and enable an analysis of, data to drive innovation while also providing privacy and security. However, adoption of PPDSA technologies has been slow because of challenges related to inadequate understanding of privacy risks and harms, limited access to technical expertise, trust, transparency among participants with regard to data collection and use, uncertainty about legal compliance, financial cost, and the usability and technical maturity of solutions.



Almost 70% of enterprises are able to classify only 50% or less of their sensitive data

Thales 2024 Data Threat Report, Global Edition

The imperative for PETs

With their ability to protect individual privacy and to eliminate the risk of data breach and deprecation of IP, Privacy Enhancing Technologies are critical in enabling organizations to leverage the deluge of data accessible to them. The question isn't whether they should adopt them, but rather, can they afford not to?

WORLD ECONOMIC FORUM

PPDSA technologies have enormous potential, but their benefit is tied to how they are developed and used. Existing confidentiality and privacy laws and policies provide important protections to individuals and communities, and attention is needed to determine how to uphold these protections using PPDSA technologies and maintain commitments to equity, transparency, and accountability. Consideration of how individuals may control the collection, linking, and use of their data should also factor into the design and use of PPDSA technologies.

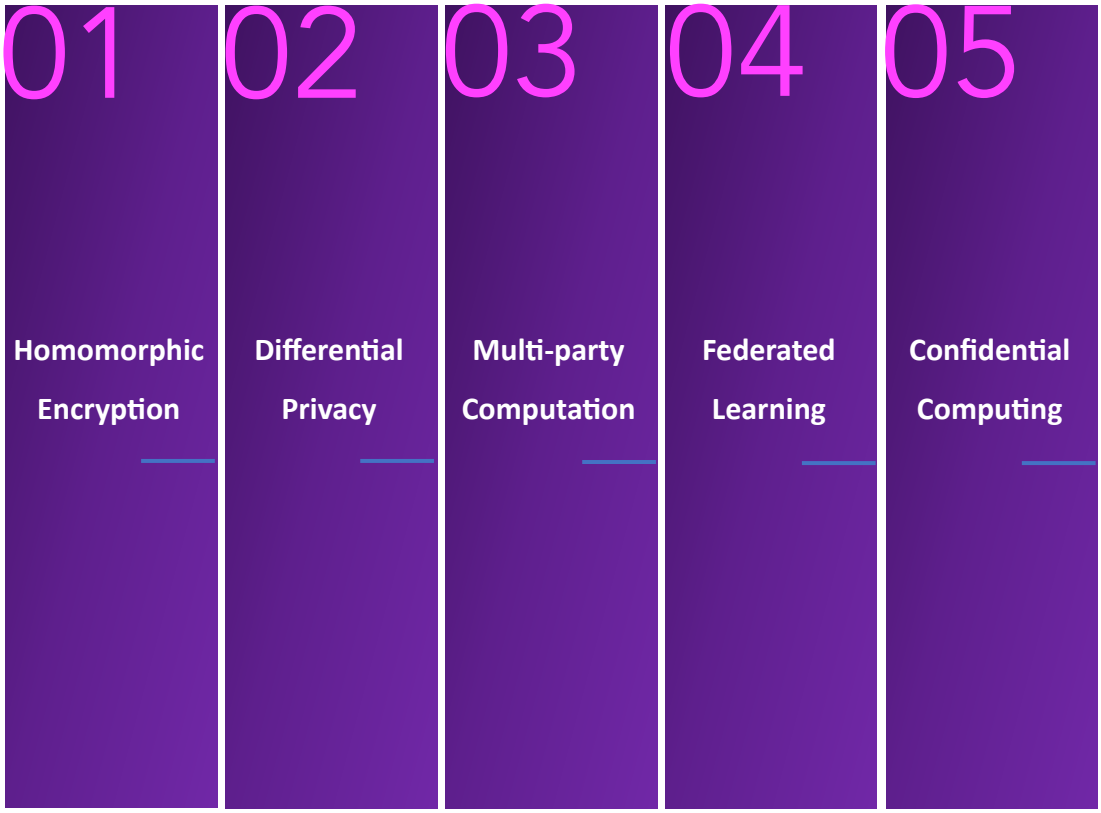
Introduction

Data drives scientific and technological breakthroughs, underpin policymaking, and power the global economy. Clinicians use data to identify the best treatments for their patients, farmers use data to predict and improve farm yields, researchers use data to generate new knowledge about natural and social phenomena, and public servants use data to create evidence-based policies. Artificial Intelligence (AI) and other emerging analytics techniques are amplifying the power of data, making it easier to discover new patterns and insights ranging from better prediction models to understand and mitigate the impacts of climate change to new methods for detecting financial crime. Although data enable science, innovation, and insights, balancing the benefits of these data-derived insights with the imperatives of privacy, security, and other values is a longstanding challenge. For example, when developing new treatment options, medical researchers may benefit from broad access to electronic health records. However, those records may contain personal health information related to individual patients, compromising the privacy and safety of those patients as well as rights under health privacy laws and regulations on the protection of human subjects. Similarly, when researchers access authorized data without safeguards on how they access the data, privacy-sensitive information such as their location or the specific type of information they are accessing may be revealed. In many domains, collaborations that could improve AI model training and accelerate progress must be balanced with ethical and legal privacy concerns and intellectual property protection concerns.

Privacy-preserving data sharing and analytics (PPDSA) solutions include technical and sociotechnical approaches that employ certain types of privacy-enhancing technologies (PETs). This generates value from, and enable, analytics on data while protecting privacy and security. Some PPDSA approaches allow users (e.g., researchers and physicians) to gain insights from sensitive data without exposing the original data itself or allow them to access shared data without being tracked or identified. Other PPDSA approaches enable data sharing by obscuring personal data or making

synthetic reflections of the original data that preserve the properties of interest in the data while protecting individual privacy.

The ever-growing volume of data presents both opportunities and challenges. Collaborative data analysis across institutions holds immense potential for scientific discovery, business innovation, and improved public services. However, sharing sensitive data often raises privacy concerns. This paper delves into the concept of secure and privacy-preserving data sharing, exploring four key cryptographic techniques that empower this approach: Homomorphic Encryption, Differential Privacy, Multi-Party Computation (MPC), and Confidential Computing.



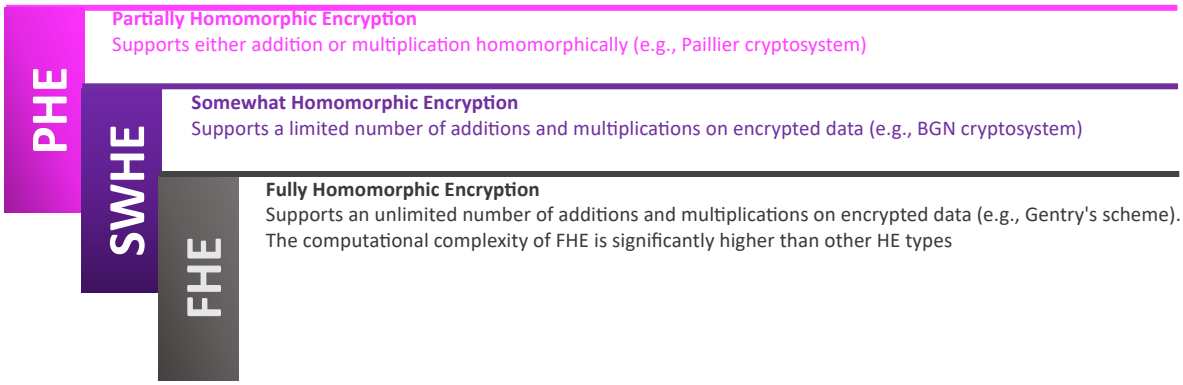
01

Homomorphic Encryption: Secure Computations on Encrypted Data

Homomorphic Encryption (HE) allows computations to be performed directly on encrypted data. Data is transformed into a ciphertext, and computations are carried out on this ciphertext. The result, also in ciphertext form, can be decrypted to reveal the outcome of the computation without ever decrypting the original data. This is akin to performing calculations on a locked box containing the data, with the result being the only information revealed. It ensures sensitive information remains confidential while enabling valuable insights to be extracted.

Technical Specifics

HE comes in various flavours, each offering different capabilities and trade-offs.



Encryption and Decryption: HE schemes involve a public key for encryption and a private key for decryption. Anyone can encrypt data using the public key, but only the authorized party with the private key can decrypt the result.

Homomorphic Operations: HE enables performing specific mathematical operations on encrypted data and obtaining the encrypted result. For example, adding encrypted numbers results in the encryption of their sum, and multiplying encrypted numbers results in the encryption of their product.

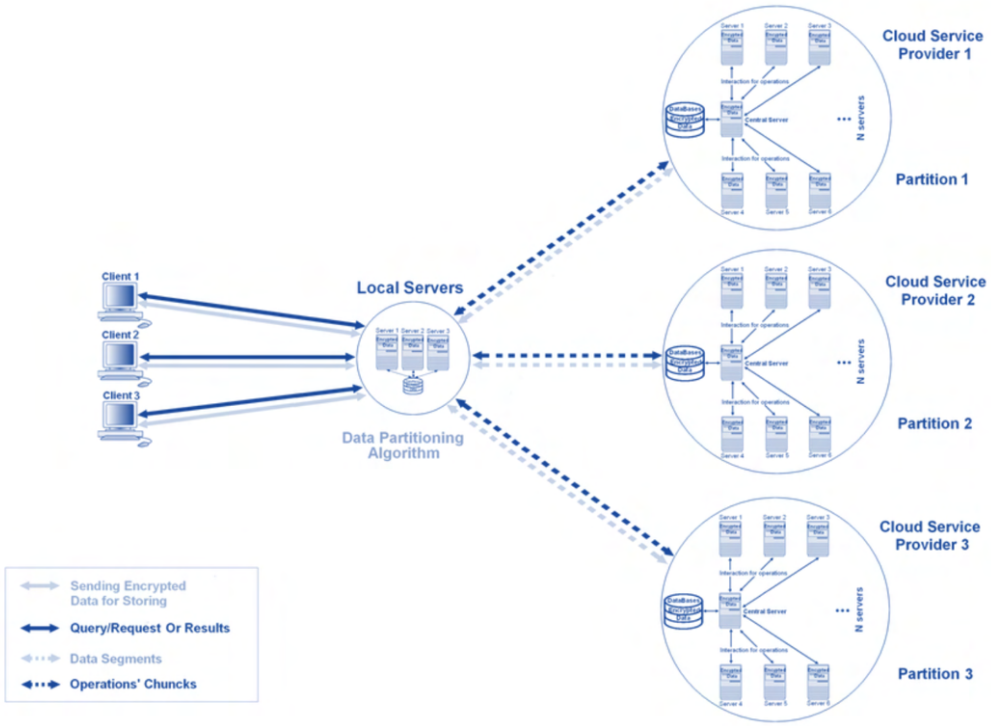


Figure 1 The proposed architecture to secure data using homomorphic encryption

Use Cases in Secure Data Sharing

Homomorphic encryption empowers various secure data sharing scenarios:

Medical Research: Collaboratively analyse patient data for research purposes without compromising individual privacy. Encrypted medical records can be analysed to identify patterns and trends while protecting sensitive patient information.

Financial Analysis: Conduct secure financial analysis on encrypted financial data. Banks or financial institutions can share encrypted financial data with analysts for risk assessment or fraud detection without revealing sensitive details.

Cloud-based Data Analysis: Upload and analyse sensitive data on cloud platforms without decryption. Businesses can leverage the scalability of cloud computing for data analysis while ensuring data confidentiality.

Government Data Sharing: Share sensitive government data for collaborative investigations or statistical analysis while maintaining individual privacy. Encrypted census data or crime statistics can be analysed without revealing details of specific individuals.

Technical Challenges and Advancements

Despite its potential, Homomorphic encryption faces some technical challenges

	Challenges	Advancements
Computational Complexity	Homomorphic computations can be computationally expensive, leading to slower processing times compared to traditional methods.	Research focuses on optimizing HE schemes and utilizing specialized hardware accelerators to improve performance.
Bootstrapping	In SWHE schemes, the number of homomorphic operations is limited. A process called bootstrapping refreshes the ciphertext to allow further computations, but it can be computationally expensive.	New bootstrapping techniques are being developed to reduce the computational overhead associated with this process.
Limited Functionality	Current FHE schemes might not support all desired operations or data types.	Ongoing research explores expanding the capabilities of FHE schemes to handle more complex computations and data structures.

Conclusion

Homomorphic encryption offers a promising approach for secure and privacy-preserving data sharing. While technical challenges remain, advancements in optimization and new research directions hold promise for a future where HE empowers secure data collaboration across various sectors. As HE continues to evolve, it has the potential to revolutionize how we share and analyse sensitive data, fostering innovation while safeguarding privacy.

02

Differential Privacy: Balancing Utility and Privacy in Data Sharing

Differential privacy (DP) is a powerful technique for sharing statistical information about datasets while protecting the privacy of individual records. It achieves this balance by carefully injecting controlled noise into the data before sharing it.

Technical Specifics

Differential privacy offers a mathematical framework that guarantees a level of privacy for individuals in a dataset. Here's how it works:

Privacy Guarantee: DP ensures that an observer cannot tell whether a specific individual's data was included in the dataset or not, by analysing the released information. The privacy guarantee is quantified by two parameters: epsilon (ε) and delta (δ). Lower values of ε and δ indicate stronger privacy guarantees.

Noise Injection: DP algorithms inject noise into the data, such as adding or subtracting random values from statistical counts. This noise masks the contribution of any individual record, making it difficult to infer information about specific individuals.

Mechanisms: Several DP mechanisms can be used to achieve differential privacy, such as the Laplace mechanism and the exponential mechanism. These mechanisms introduce noise in different ways to achieve the desired level of privacy and data utility.

Share user data with advertisers in a privacy-preserving manner. DP ensures user profiles remain confidential while enabling targeted advertising campaigns based on aggregated user behaviour.

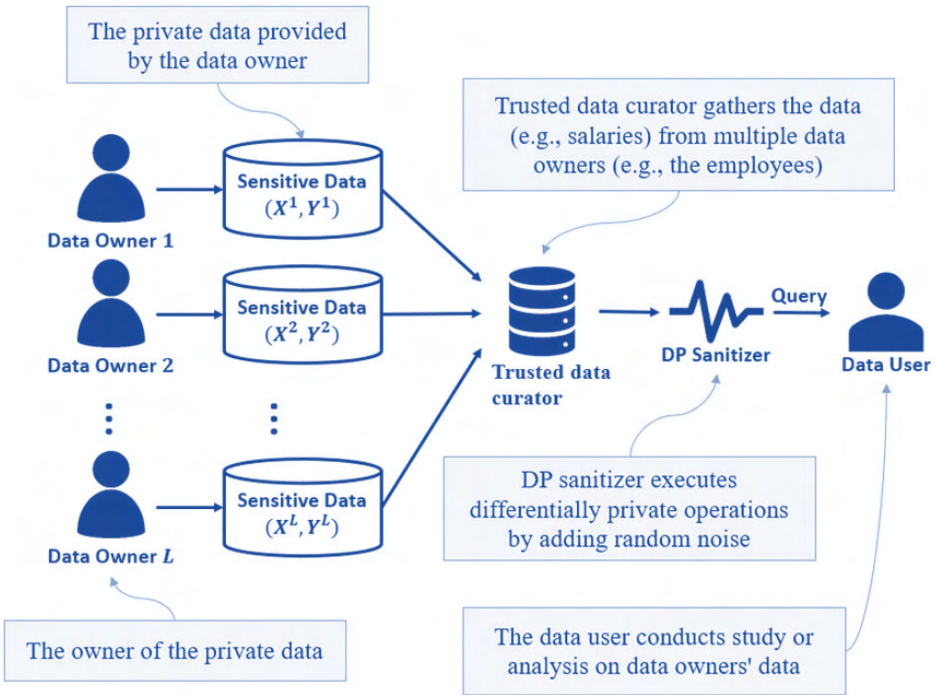


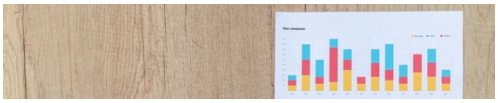
Figure 2 Working Methodology for Differential Privacy

Use Cases in Secure Data Sharing

Examples of how differential privacy empowers data sharing:

Publishing Census Data: Release aggregated statistics about a population while protecting individual privacy. DP ensures that demographic information remains confidential, even for small populations.

Medical Research: Analyse anonymized medical records for research purposes without compromising patient identities. DP allows researchers to extract valuable insights from large datasets while protecting individual patient information.



Targeted Advertising: Share user data with advertisers in a privacy-preserving manner. DP ensures user profiles remain confidential while enabling targeted advertising campaigns based on aggregated user behaviour.

Machine Learning with Sensitive Data: Train machine learning models on sensitive data while protecting individual records. DP helps ensure the model learns from broader patterns without revealing details about specific data points used for training.

Technical Challenges and Advancements

Differential Privacy offers many benefits but faces some technical challenges:

	Challenges	Advancements
Privacy-Utility Trade-off	Adding more noise to achieve stronger privacy guarantees can also reduce the accuracy and usefulness of the data (utility).	Researchers are exploring new DP mechanisms and algorithms that offer better trade-offs between privacy and utility. Composition theorems are being developed to handle the composition problem while minimizing noise accumulation. Synthetic data generation techniques are being investigated to create realistic anonymized datasets that can be used in conjunction with DP for small datasets.
Composition Problem	Combining multiple differentially private operations can amplify the noise and significantly reduce data utility.	
Data Utility for Small Datasets	DP can be less effective for very small datasets as the noise injection can significantly distort the underlying data.	

Conclusion

Differential privacy provides a valuable approach for secure data sharing by enabling the release of statistical insights without compromising individual privacy. While challenges exist in balancing privacy and utility, ongoing research efforts aim to improve DP techniques and address its limitations. By fostering innovation in DP algorithms and addressing the trade-offs involved, this technique holds immense potential for a future where data sharing benefits society while upholding individual privacy rights.

03

Multi-Party Computation: Collaborative Analysis Without Sharing Data

Multi-party computation (MPC) empowers multiple parties to jointly compute a function on their private data inputs without revealing those inputs to each other.

Imagine several parties holding pieces of a puzzle; through MPC, they can collectively solve the puzzle without revealing their individual pieces. This fosters secure data collaboration while ensuring each party retains control over its own sensitive data.

Technical Specifics

MPC relies on complex cryptographic protocols that allow secure communication and computation between parties without revealing the underlying data.

Here's a breakdown of the key concepts:

Secret Sharing: Data is divided into secret shares and distributed among the participating parties. No single party possesses the complete data; only by combining their shares can they reconstruct the final result.

Garbled Circuits: In this approach, the function to be computed is transformed into a garbled circuit. Each gate in the circuit is replaced with a set of encrypted instructions. Parties evaluate these instructions based on their secret shares without revealing the actual data.

Secure Comparison: MPC protocols often involve secure comparison techniques that allow parties to compare their inputs without revealing the actual values. This is crucial for functionalities like joint risk assessment or collaborative filtering.

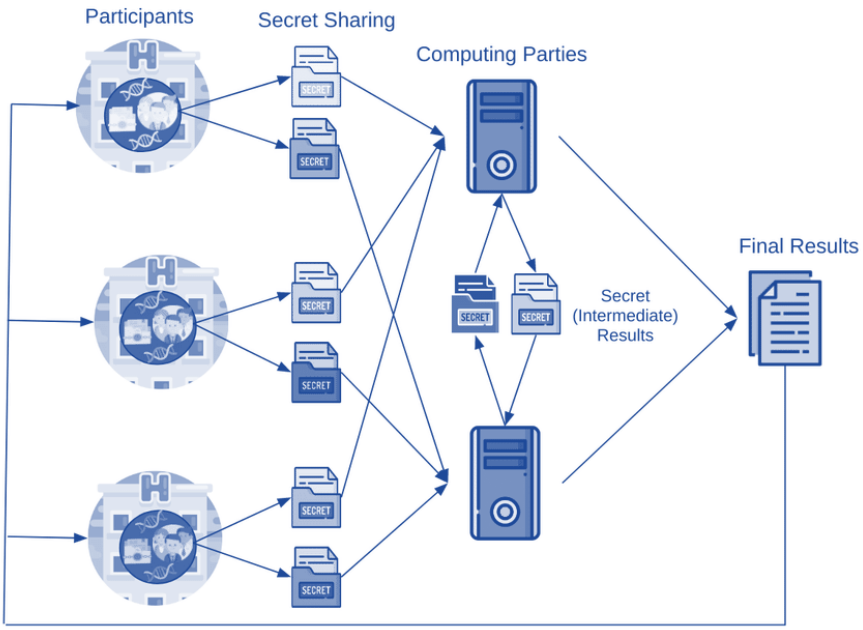


Figure 3 Secure Multi-Party Computation

Use Cases in Secure Data Sharing

MPC enables secure collaboration in various scenarios:

Financial Analysis: Financial Analysis: Banks can jointly assess creditworthiness of loan applicants without sharing individual customer data. This fosters collaboration while protecting sensitive financial information.

Fraud Detection: Credit card companies can combine data to identify fraudulent transactions without revealing individual customer details. Collaborative analysis helps identify patterns and prevent fraud more effectively.

Scientific Research: Researchers from different institutions can analyse sensitive data sets (e.g., genomic data) for joint research projects without sharing the raw data. This allows for scientific advancements while protecting participant privacy.

Joint Auctions: Bidding companies can participate in secure auctions without revealing their individual bidding strategies. MPC ensures fair competition while protecting sensitive pricing information.

Technical Challenges and Advancements

Despite its potential, MPC faces some technical hurdles:

	Challenges`	Advancements
Computational Complexity	MPC computations can be computationally expensive, especially for complex functions and large datasets. This can impact the efficiency and scalability of the protocol.	Researchers are actively exploring new protocols and techniques to improve the efficiency and scalability of MPC. This includes utilizing hardware accelerators and optimizing communication channels. Additionally, research is ongoing into minimizing the reliance on trusted parties in MPC protocols.
Trusted Parties	Some MPC protocols rely on trusted parties to facilitate communication and ensure protocol execution. This introduces a potential trust dependency.	
Scalability	Scaling MPC to a large number of participants can be challenging due to increased communication overhead and computational complexity.	

Conclusion

Multi-party computation offers a powerful approach for secure data collaboration. It allows various parties to share insights without compromising their sensitive data. While challenges remain in terms of efficiency and scalability, ongoing research holds promise for a future where MPC becomes a practical tool for secure and privacy-preserving data sharing across diverse applications. By overcoming these challenges, MPC can empower collaborative data analysis and unlock new possibilities for innovation in various sectors.

04

Federated Learning: Secure and Private Data Sharing for Machine Learning

Federated learning is a machine learning approach that enables collaborative training of models without directly sharing the underlying data. This is particularly beneficial when dealing with sensitive data where privacy is a major concern.

Technical Specifics

Distributed Training: Data remains on individual devices or servers (called clients) where local models are trained.

Model Updates: Clients only share model updates (changes in weights and biases) with a central server (coordinator). These updates contain no raw data.

Aggregation: The coordinator aggregates the received updates to improve a global model.

Privacy-Preserving Techniques: Techniques like differential privacy can be used to add noise to the updates, further protecting individual contributions.

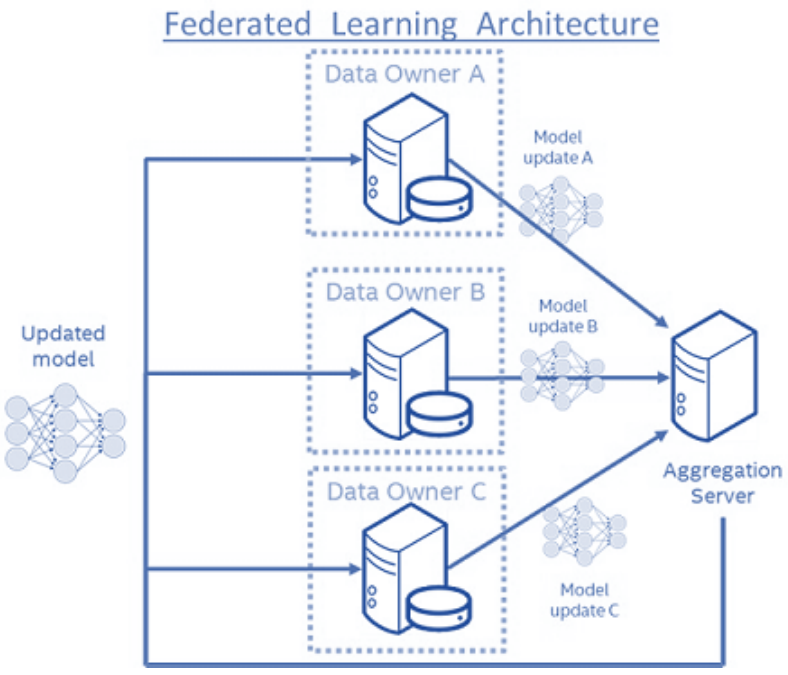


Figure 4 Basic Federated Learning Architecture

Use Cases in Secure Data Sharing

Federated learning use cases include:-

Mobile Keyboard Prediction: Smartphone keyboards can learn personalized suggestions by federated learning on user typing data stored on the devices.

Medical Diagnosis: Hospitals can train a disease prediction model on their patient data without sharing the sensitive medical information itself.

Financial Fraud Detection: Banks can collaboratively build fraud detection models without revealing individual customer transactions.

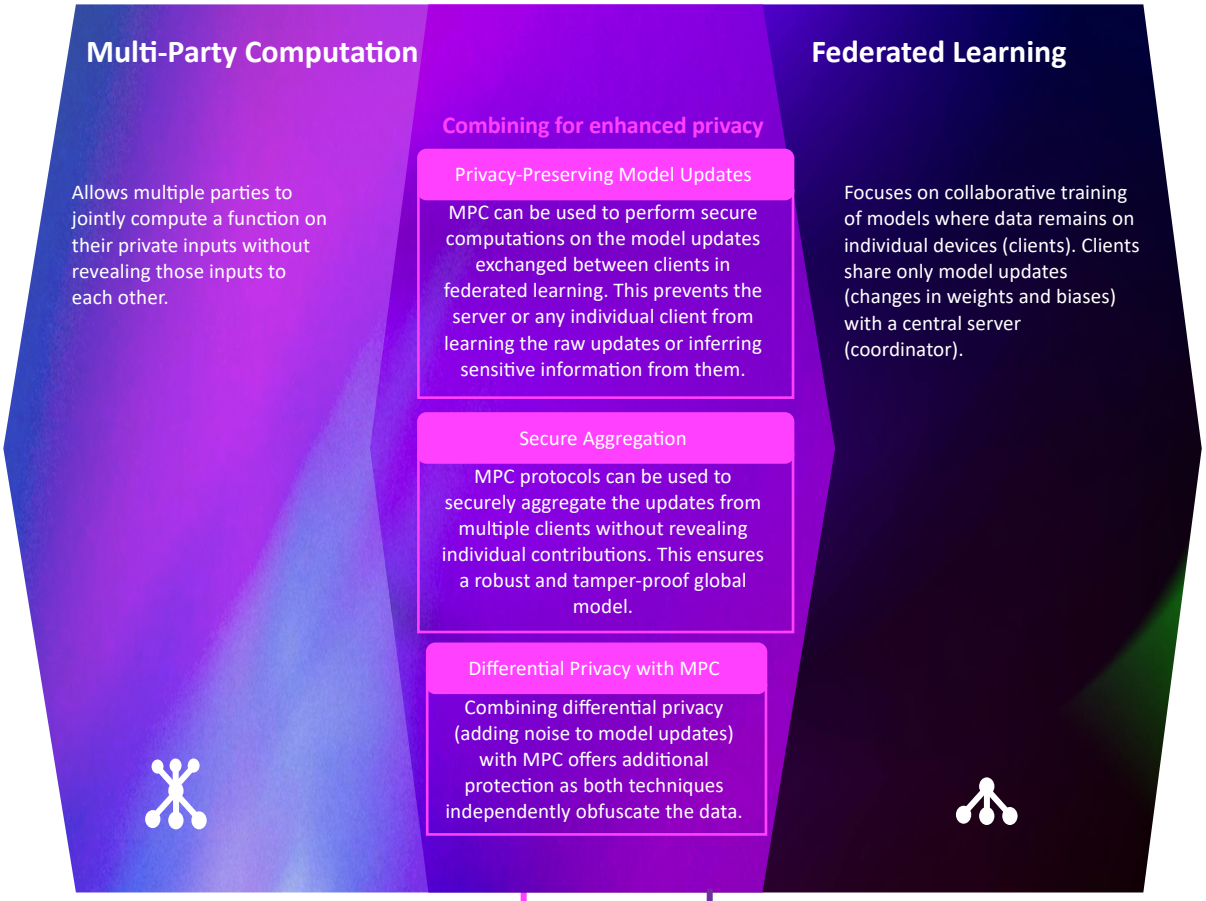
Technical Challenges and Advancements

	Challenges	Advancements
Communication Overhead	Frequent communication of model updates can be resource-intensive, especially for geographically distributed clients.	Efficient Aggregation Protocols: Techniques, such as selective aggregation, or model averaging, can reduce communication overhead.
Non-IID Data	Clients might have data with different distributions (non-IID), hindering the effectiveness of the global model.	Federated Transfer Learning: Pre-training a base model on a diverse dataset helps address non-IID data issues.
Privacy Leakage	Even with model updates, there's a risk of inferring sensitive information through reconstruction attacks.	Differential Privacy with Secure Aggregation: Adding controlled noise to model updates combined with secure aggregation methods can further enhance privacy protection.

Federated learning is a rapidly evolving field with ongoing research to address these challenges and improve its effectiveness. By enabling collaborative learning without compromising privacy, it holds significant promise for various applications that rely on sensitive data.

Federated Learning and Multi-Party Computation together

Federated learning and multi-party computation (MPC) are complementary techniques that can be combined to achieve even stronger privacy guarantees in secure data sharing for machine learning. Here's how they work together:



BENEFITS

Stronger Privacy Guarantees
MPC offers a higher level of privacy compared to just adding noise in federated learning. It protects against potential reconstruction attacks where an adversary could try to infer individual data points from the aggregated model.

Improved Trust
By eliminating the need for clients to trust the server with their updates, MPC fosters a more secure collaborative environment.

CHALLENGES

Computational Overhead
MPC computations can be computationally expensive, impacting training speed, especially for complex models.

Scalability
MPC protocols might not scale efficiently to a large number of clients, limiting its applicability in certain scenarios.

Research Directions:

- Development of more efficient MPC protocols specifically designed for federated learning tasks.
- Exploring alternative privacy-preserving techniques that can be integrated with federated learning.

Overall, combining federated learning with MPC offers a powerful approach for collaborative machine learning while ensuring strong privacy guarantees for sensitive data. As research progresses, these techniques are expected to play a significant role in unlocking the potential of data sharing in a privacy-conscious manner.

Federated Learning Transforming Smart Cities

Federated learning can solve challenges in modern digitised transportation systems, such as data privacy, calculation processing, and communication delay. Smart city programs are critical infrastructure requiring robust cybersecurity based on privacy enhancing technologies

05

Confidential Computing: Secure Enclaves for Privacy-Preserving Data Processing

Confidential computing emerges as a game-changer for secure privacy-preserving data sharing. It leverages hardware-based security features within processors to create trusted execution environments (TEEs) where data remains encrypted even during processing. This ensures sensitive data processing occurs in a secure isolation layer, safeguarding privacy while facilitating valuable data analysis.

Technical Specifics

Confidential computing relies on hardware capabilities within modern processors:

- **Trusted Execution Environments (TEEs):** These are isolated execution environments within the processor that protect data and code confidentiality even from the main operating system and any potential attackers. They provide a secure enclave for processing sensitive data.
- **Secure Enclave Management:** Software tools and libraries manage the creation, provisioning, and access controls for TEEs. This ensures authorized applications can leverage the secure enclave for data processing.
- **Encryption and Decryption:** Data is encrypted before entering the TEE and decrypted only after processing is complete. End-to-end Data Protection (E2EDP) ensures sensitive data remains protected throughout the entire processing lifecycle within the enclave.

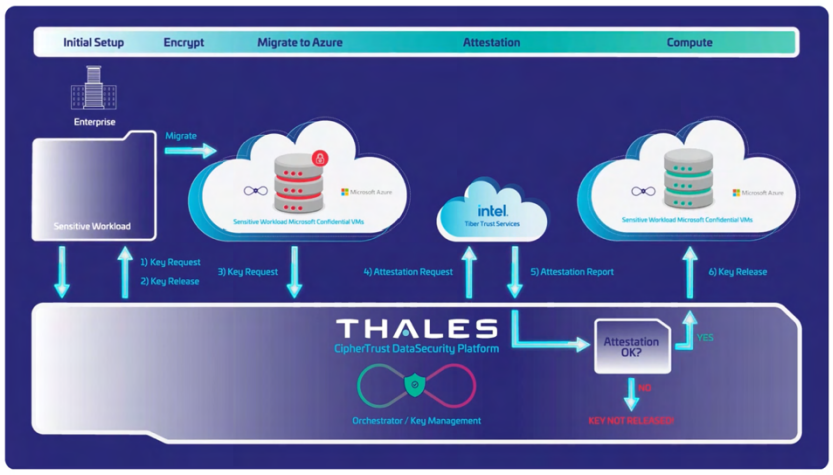


Figure 5 Example Model for Confidential Computing: Thales End-to-end Data Protection

Use Cases in Secure Data Sharing

Confidential computing empowers secure data analysis in various scenarios:

Cloud-based Data Analytics: Businesses can leverage the scalability of cloud computing for sensitive data analysis without compromising privacy. Confidential computing ensures data remains encrypted even on shared cloud infrastructure.

Financial Transactions: Enhance security for financial transactions like clearing and settlement processes. Confidential computing protects sensitive financial data during processing within the TEE.

Supply Chain Management: Securely analyse and share sensitive supply chain data (e.g., pricing information, intellectual property) across collaborating parties. Confidential computing safeguards data confidentiality during collaborative analysis in a TEE.

Healthcare Data Analysis: Collaboratively analyses encrypted patient data for research purposes while maintaining patient privacy. Confidential computing ensures sensitive medical data remains protected within the secure enclave.

Technical Challenges and Advancements

Confidential computing considerations

	Challenges	Advancements
Limited Adoption	TEE technology is evolving. Widespread adoption requires broader hardware and software support. Integrating confidential computing with existing applications can be complex.	Software development frameworks are being developed to simplify application integration with confidential computing.
Performance Overhead	Processing data within the TEE can introduce some performance overhead compared to traditional processing methods.	Processor manufacturers are actively improving TEE functionalities and performance.
Standardization	Standardizing interfaces and APIs for confidential computing across different hardware platforms is crucial for broader adoption and interoperability.	Industry collaboration is fostering standardization efforts to ensure compatibility across platforms.

Conclusion

Confidential computing offers a promising approach for secure privacy-preserving data processing. By leveraging hardware-based security enclaves, it empowers data collaboration without compromising privacy. While challenges regarding adoption, performance, and standardization remain, ongoing advancements hold significant promise for a future where confidential computing becomes a cornerstone of secure and privacy-conscious data utilization across various sectors. As the technology matures and limitations are addressed, confidential computing has the potential to revolutionize the way we analyse and share sensitive data, fostering secure collaboration and innovation.

Confidential Computing used at scale within Central Bank Digital Currency.

"A CBDC designed on confidential computing principles can be as private as cash is today, and still allow tracking of aggregate economic trends and preventing money-laundering-enabled crime."

Forbes 2023

Digital Assets and Privacy Preserving Technologies

Digital Asset and specifically CBDC design choices will significantly impact user privacy. Some designs concentrate a lot of data, including user identities and transaction details, with the central bank. This raises concerns about mass surveillance and potential misuse of this data.

Privacy-preserving technologies hold immense potential in the digital asset space, addressing critical concerns around user privacy and security without hindering innovation.

Here are some key use cases:

1. Secure Transactions:

Zero-knowledge proofs: Allow users to prove they possess certain digital assets (e.g., cryptocurrency) without revealing the exact amount or transaction details. This safeguards financial privacy while ensuring transaction validity.

2. Enhanced Regulatory Compliance:

Homomorphic encryption: Enables regulatory bodies to analyse anonymized blockchain data for anti-money laundering (AML) and Know Your Customer (KYC) purposes without compromising user privacy. This fosters a balance between transparency and user protection.

3. Confidential Smart Contracts:

Secure multi-party computation (MPC): Allows multiple parties to execute smart contracts on a blockchain without revealing any private data involved in the contract's logic. This facilitates secure and private business transactions on blockchains.

4. Decentralized Identity Management:

Self-sovereign identity (SSI): Empowers users to control their digital identities on blockchains. Users can choose what information to share with different entities, promoting user control over personal data.

5. Secure Cryptographic Wallets:

Homomorphic encryption: Allows users to perform basic operations (e.g., balance checks) on their cryptocurrency holdings within the wallet without decrypting the entire private key. This enhances security by minimizing exposure of sensitive cryptographic information.

6. Secure DeFi (Decentralized Finance) Applications:

Zero-knowledge proofs: Users can prove their eligibility for DeFi services (e.g., loans) without revealing their entire financial data. This fosters broader participation in DeFi while safeguarding user privacy.

7. Privacy-Preserving Analytics:

Federated learning: Allows different parties to collaboratively train machine learning models on their private datasets without sharing the underlying data itself. This enables valuable data analysis for fraud detection or market insights while protecting user privacy.

8. Secure Data Sharing for Regulatory Reporting:

MPC: Financial institutions can share sensitive data required for regulatory reporting with auditors or regulators in a privacy-preserving manner. This ensures compliance while safeguarding sensitive financial information.

Overall, privacy-preserving technologies play a crucial role in the future of digital assets. By enabling secure transactions, fostering regulatory compliance, and empowering user control over data, these technologies pave the way for a more secure, transparent, and user-centric digital asset ecosystem.

Digital Sovereignty and Privacy Preserving Technologies

Privacy-preserving secure data sharing plays a critical role in achieving Digital Sovereignty for several reasons:

1. Enables Data Sharing While Maintaining Control:

Data as a National Asset: In the digital age, data is a valuable national asset. Digital Sovereignty emphasizes a nation's control over its data. Privacy-preserving techniques allow data sharing for specific purposes (e.g., research, public policy) without compromising the privacy of individuals or revealing sensitive information.

Collaboration without Compromising Privacy: Countries can collaborate on projects that require data sharing, such as scientific research or public health initiatives, while ensuring individual privacy is protected. This fosters international cooperation without sacrificing data control.

2. Strengthens Data Security and Trust:

Reduced Risk of Data Breaches: Privacy-preserving techniques like homomorphic encryption or secure multi-party computation keep data encrypted even during analysis. This minimizes the risk of data breaches and unauthorized access, protecting sensitive information.

Increased Public Trust: By demonstrating a commitment to data privacy through secure sharing methods, governments can build trust with citizens. This is essential for encouraging data sharing and participation in digital initiatives.

3. Supports Data-Driven Decision Making:

Access to Diverse Data Sets: Privacy-preserving techniques allow access to data sets from various sources while protecting individual privacy. This enables governments to make informed decisions based on a wider range of data, leading to more effective policies.

Empowers Individuals: Individuals can choose to share their data for specific purposes with the assurance of privacy protection. This empowers citizens to contribute to data-driven decision making without compromising their own privacy.

4. Fosters Innovation in Secure Data Ecosystems:

Development of New Technologies: The need for secure data sharing drives innovation in privacy-preserving techniques. This fosters the development of new technologies that can be beneficial for various sectors beyond government, like healthcare and finance.

Creates a Secure Digital Infrastructure: By promoting secure data sharing practices, governments can contribute to building a more secure digital infrastructure. This benefits all stakeholders within the digital ecosystem.

Overall, privacy-preserving secure data sharing is a crucial element of Digital Sovereignty. It empowers nations to control their data, collaborate securely, and leverage data for the benefit of their citizens while ensuring individual privacy remains a top priority.

The Role of Hardware Security Modules (HSMs)

Hardware security modules (HSMs) and privacy-preserving techniques (PPTs) e.g. homomorphic encryption, multi-party computation (MPC), federated learning, confidential computing, and differential privacy offer complementary functionalities for a secure privacy-preserving data sharing platform. Here's how they can work together:

Hardware Security Modules (HSMs):

- Key Management and Secure Execution: HSMs provide a secure environment for storing and managing cryptographic keys used by PPTs as part of a Key Management Service (KMS). These keys are essential for encryption and decryption operations within PPTs.
- Tamper Evident and Secure Processing: HSMs are tamper-evident devices, meaning any attempt to tamper with them will be detected. This adds an extra layer of security to the platform, ensuring the integrity of the data and the computations performed on it.
- Offloading Processing: Some PPTs, particularly homomorphic encryption, can be computationally expensive. HSMs can offload these heavy computations from the main platform, improving performance and scalability.

Synergy: Privacy-Preserving Techniques (PPTs) and HSM

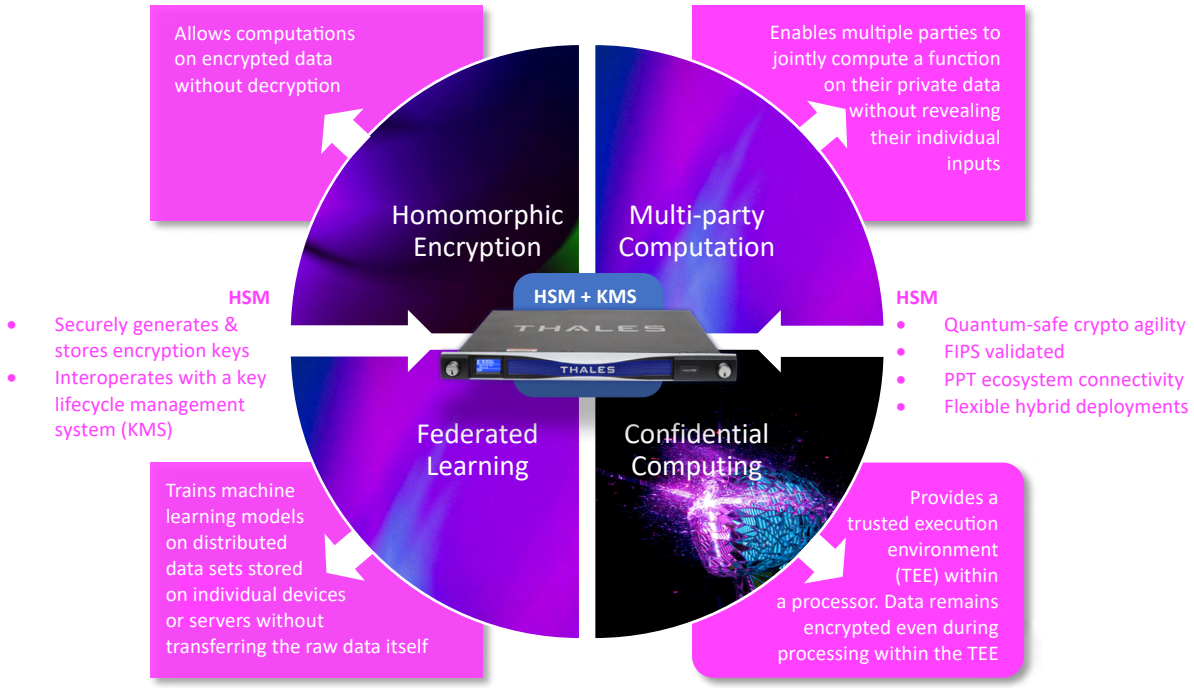


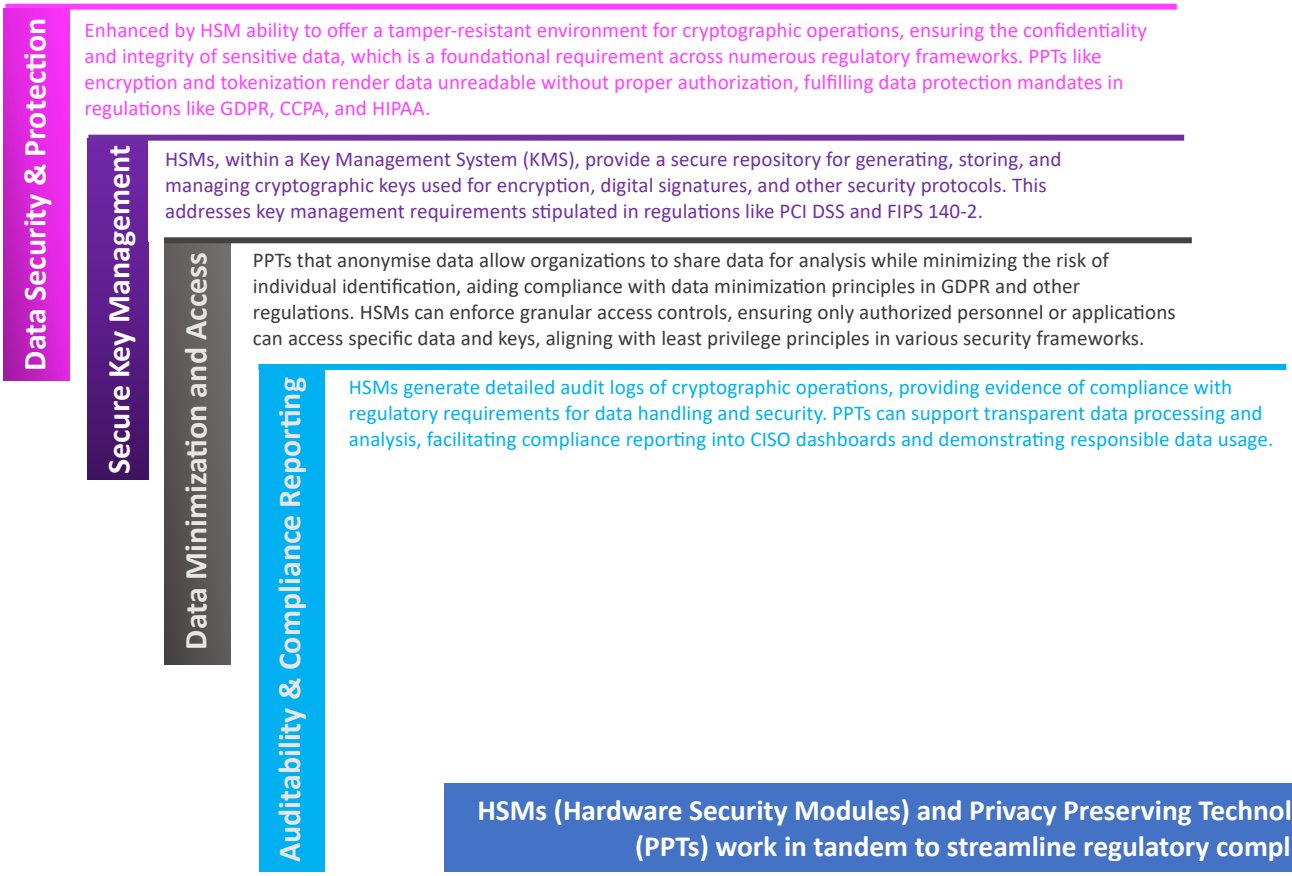
Fig. 6 HSM support of PPTs. Example show is HSM manufactured by Thales

Differential Privacy: Adds controlled noise to data to protect individual privacy while allowing for statistical analysis. HSMs are not directly involved in differential privacy

By combining HSMs with PPTs, the platform can achieve a high level of security and privacy for data sharing:

- HSMs provide a secure foundation: Safeguarding cryptographic keys and perform secure computations, ensuring the data integrity and confidentiality throughout the platform.
- PPTs offer specific privacy guarantees: Each PPT caters to different privacy needs. Choosing the right combination of PPTs based on the specific use case ensures data remains private while allowing for collaborative analysis and insights.

HSMs, PPTs and Easing Regulatory Compliance



Regulatory Compliance Examples

Healthcare: Encryption enforced by HSMs protects patient health information (PHI) as required by HIPAA. De-identification and anonymization techniques facilitate research while adhering to privacy regulations.

Government and Defence: HSMs safeguard classified information, aligning with stringent government and defence security standards. Securing communication channels in the face of quantum computer-based attacks

Financial Services: HSMs protect sensitive financial data and transaction keys, enabling compliance with PCI DSS requirements. Tokenization helps safeguard cardholder data, further reducing compliance scope. Here is an overview of a sample global regulatory compliance

The Securities and Futures Commission conclusions on virtual asset trading platforms 2023-05

Hardware Security Modules (HSMs) play a crucial role in maintaining the security and integrity of cryptographic operations, particularly in the financial and regulatory contexts referenced in the SFC (Securities and Futures Commission) consultation conclusion document. In the context of this consultation, which discusses enhancing security frameworks within financial institutions, HSMs are vital for several reasons:

Secure Key Management	HSMs ensure that cryptographic keys are not exposed or tampered with, even in a highly regulated environment
Encryption and Decryption Operations	HSMs perform encryption and decryption operations directly within the secure hardware, minimizing the risk of data leakage or unauthorized access
Regulatory Compliance	HSMs are often mandatory, e.g. in GDPR, DORA, PDCI DSS, because they provide a certified level of security
Privacy-Preserving Technologies Integration	Integration with, for example, homomorphic encryption, or secure Multi-Party Computation as discussed earlier, enables specific new use cases and business value. In many cases can lead to the creation of new innovative digital financial products
Digital Signatures and Authentication	HSMs are also used for generating and verifying digital signatures, which are crucial for ensuring the authenticity and integrity of transactions. This supports the secure and compliant operations required in financial regulations. In the context of privacy-preserving technologies, HSMs can securely handle the signing processes, ensuring that data remains tamper-proof and authentic
Secure Key Management	HSMs ensure that cryptographic keys are not exposed or tampered with, even in a highly regulated environment

SFC Conclusion: By integrating HSMs within their infrastructure, financial institutions can enhance data security, meet regulatory requirements, and support privacy-preserving technologies; critical elements discussed in the SFC consultation conclusion document.

Hong Kong Monetary Authority (HKMA) Provision of Custodial Services 2024-02-20

The Hong Kong Monetary Authority (HKMA) outlines the regulatory requirements and guidelines for managing risks associated with technology in the financial sector. Within this context, Hardware Security Modules (HSMs) are essential tools for ensuring secure cryptographic operations, which are critical for protecting sensitive financial data and maintaining compliance with regulatory standards.

Secure Cryptographic Operations	HSM tamper-resistant hardware, provide a secure environment for performing cryptographic operations such as encryption, decryption, key management, and digital signatures. These operations are crucial for safeguarding sensitive data from unauthorized access, and breach resistance, especially in financial transactions
Regulatory Compliance	The use of HSMs supports compliance with data protection regulations, such as the Personal Data (Privacy) Ordinance (PDPO) in Hong Kong. These regulations require organizations to implement strong measures to protect personal data. HSMs ensure that the cryptographic processes involved in these measures are secure HSMs also support the implementation of strong authentication mechanisms, which are critical for meeting the HKMA's requirements for identity verification and access control
Support for Privacy-Preserving Technologies	Privacy-preserving technologies, such as homomorphic encryption and zero-knowledge proofs, require robust key management and secure cryptographic operations to function effectively. HSMs provide the necessary secure environment for generating, storing, and using cryptographic keys in these privacy-preserving techniques By integrating HSMs with privacy-preserving technologies, financial institutions can ensure that sensitive data is protected not only at rest but also during processing and transmission
Linking HSMs to Privacy-Preserving Technologies	Encryption and Data Masking: HSMs play a critical role in encryption and data masking, which are foundational to privacy-preserving technologies. These technologies allow financial institutions to process data without exposing it, thus protecting customer privacy even during analytics or sharing of information

HKMA Conclusion: In summary, HSMs are integral to meeting the HKMA's regulatory requirements for secure cryptographic operations and data protection. They also facilitate the implementation of privacy-preserving technologies by ensuring that cryptographic keys and operations are securely managed, thus protecting sensitive data in compliance with regulatory standards.

Hong Kong Monetary Authority (HKMA) DLT Risk management 2024-04-16

In the context of the Hong Kong Monetary Authority's (HKMA) guidelines on Distributed Ledger Technology (DLT) Risk Management dated April 16, 2024, Hardware Security Modules (HSMs) are critical for ensuring the security and integrity of cryptographic operations that underpin DLT systems. Here's how HSMs are used and their connection to privacy-preserving technologies:

Secure Key Management in DLT Systems	HSMs provide a secure environment for managing cryptographic keys that are essential in DLT systems. In DLT, cryptographic keys are used for signing transactions, validating blocks, and managing access to the ledger. The HKMA guidelines emphasize the importance of securing these keys to prevent unauthorized access or tampering, which could compromise the entire DLT system HSMs ensure that private keys used for signing transactions are securely stored and processed, reducing the risk of key theft or misuse, which is a critical concern in DLT-based financial services
Transaction Integrity and Authentication:	HSMs are used to ensure the integrity and authenticity of transactions within a DLT network. By securely managing the cryptographic processes involved in transaction signing and verification, HSMs help maintain the trustworthiness of the DLT ledger, as mandated by the HKMA. Prevention of fraudulent transactions depends on this, and ensures that all actions on the ledger are traceable and legitimate
Regulatory Compliance and Risk Mitigation:	The HKMA's DLT Risk Management guidelines highlight the need for financial institutions to adhere to strict security standards and mitigate risks associated with the use of DLT. HSMs help meet these requirements by providing a high level of security for cryptographic operations, which is necessary for regulatory compliance. HSMs also facilitate the implementation of secure access controls, which are essential for protecting sensitive data and ensuring that only authorized parties can interact with the DLT system
Linking HSMs to Privacy-Preserving Technologies in DLT	**Privacy Preserving Transactions** Privacy-preserving technologies in DLT, such as zero-knowledge proofs (ZKPs) and confidential transactions, rely on secure cryptographic operations to protect the privacy of participants and transaction data. For example, in ZKPs, HSMs can securely generate and manage the cryptographic proofs that allow participants to verify the validity of a transaction without revealing sensitive details **Confidential Data Sharing** In DLT systems, privacy-preserving technologies often involve sharing confidential data across multiple parties while keeping it encrypted and secure. HSMs ensure separation of duty with encryption keys used in these processes securely managed, preventing unauthorized access to data. This is particularly important in financial services, where maintaining the confidentiality of transaction data is critical for regulatory compliance and protecting customer privacy. **Support for Advanced Cryptographic Techniques** Homomorphic encryption is one such technique that allows computations to be performed on encrypted data. In the context of DLT, this can enable privacy-preserving analytics and data processing while ensuring that sensitive information remains protected. By securely managing the cryptographic keys and processes involved in these techniques, HSMs enable financial institutions to leverage privacy-preserving technologies in a secure and compliant manner

HKMA DLT Conclusion: HSMs are integral to secure cryptographic operations. They secure the implementation of DLT systems in accordance with the HKMA's DLT Risk Management guidelines, ensuring the integrity, confidentiality, and authenticity of transactions on the ledger. Additionally, HSMs enable the secure use of privacy-preserving technologies within DLT, allowing financial institutions to protect sensitive data while complying with regulatory requirements.

The Securities and Exchange Commission (SEC)

The Securities and Exchange Commission (SEC) has increasingly focused on the role of technology and risk management in ensuring robust regulatory compliance within the financial industry. In this context, Hardware Security Modules (HSMs) and Privacy-Preserving Technologies (PPTs) are seen as critical components that can help institutions meet stringent security and privacy requirements.

HSMs in SEC Guidelines:

Secure Cryptographic Operations	The SEC emphasizes the importance of protecting sensitive financial data and ensuring the integrity of transactions. HSMs are essential for securely managing cryptographic keys, which are used for encryption, decryption, digital signatures, and other cryptographic functions that are critical for safeguarding data. By securely storing and processing these keys, HSMs help financial institutions comply with SEC regulations that require the protection of customer data and the integrity of financial transactions
Enhanced Authentication and Access Control:	SEC guidelines often require robust authentication and access control mechanisms to prevent unauthorized access to sensitive systems and data. HSMs support the security infrastructure needed to implement these controls. This helps in mitigating risks associated with insider threats or cyberattacks

Privacy-Preserving Technologies in SEC Guidelines:

Data Privacy and Protection	Privacy-preserving technologies such as homomorphic encryption, secure multi-party computation, and zero-knowledge proofs enable financial institutions to process and share data without exposing sensitive information. The SEC recognizes the importance of these technologies in maintaining data privacy, say for GDPR or CCPA, while allowing for necessary data analysis and compliance reporting. By adopting PPTs, institutions can ensure that they comply with privacy regulations, while still fulfilling their obligations to report data to regulators
Compliance with Data Protection Regulations	The SEC's guidelines often intersect with other regulatory frameworks that prioritize data privacy. By integrating PPTs into their operations, financial institutions can more easily comply with multiple regulatory requirements, including those that mandate data minimization and secure data handling practices. PPTs allow institutions to provide the necessary transparency and reporting to regulators without compromising the privacy of their customers, thus easing the burden of regulatory compliance.
Easing SEC Regulatory Compliance	**Automation and Integration:** The SEC encourages the use of technology to automate and streamline compliance processes. HSMs and PPTs can be integrated into compliance workflows to automatically enforce data protection and security measures, reducing the likelihood of human error and ensuring consistent compliance with regulations. For example, HSMs, together with a centralised key management system can automate the encryption of data at rest and in transit, while PPTs can automate the secure sharing of data with regulators, making the compliance process more efficient and less prone to breaches. **Auditability and Transparency:** Both HSMs and PPTs enhance the auditability of financial operations. HSMs ensure that all cryptographic operations are logged and can be audited, providing clear evidence of compliance with SEC requirements. PPTs, on the other hand, ensure that data privacy is maintained throughout the audit process, allowing institutions to demonstrate compliance without exposing sensitive information. This dual focus on security and privacy helps institutions navigate complex regulatory environments and reduces the risk of non-compliance penalties.

SEC Conclusion: The SEC's guidelines foresee the use of HSMs and Privacy-Preserving Technologies as essential tools in ensuring that financial institutions can securely manage data, protect customer privacy, and maintain compliance with regulatory requirements. These technologies provide the necessary infrastructure to safeguard data and automate compliance processes, thereby easing the regulatory burden on financial institutions.

Conclusion

The combination of HSMs and PPTs offers a robust framework for addressing the diverse and often overlapping requirements of global regulatory bodies. By ensuring data security, secure key management, privacy protection, and auditable compliance, organizations can navigate the complexities of the regulatory landscape more effectively and efficiently. This not only reduces the risk of non-compliance penalties but also builds trust with customers and stakeholders, demonstrating a commitment to responsible data handling and ethical AI practices.

References

1. Dwork, C. (2006). Differential privacy. In 33rd International Colloquium on Automata, Languages and Programming, part II (ICALP 2006) (pp. 1-12). Springer Berlin Heidelberg.
2. Acar, A., Aksu, H., Uluagac, A. S., & Conti, M. (2018). A survey on homomorphic encryption schemes: Theory and implementation. ACM Computing Surveys (CSUR), 51(4), 1-35.
3. Schneier, B. (1996). Applied cryptography: protocols, algorithms, and source code in C.
4. Sharma, V., & Malhotra, J. (2013). Data Tokenization: A solution for PCI DSS Compliance. International, Journal of Computer Applications,
5. Yang, Q., Liu, Y., Chen, T., & Tong, Y. (2019). Federated machine learning: Concept and applications. ACM Transactions on Intelligent Systems and Technology (TIST), 10(2), 1-19.
6. Yao, A. C. (1982). Protocols for secure computations. In 23rd annual symposium on foundations of computer science (sfcs 1982)(pp. 160-164). IEEE
7. Goldwasser, S., Micali, S., & Rackoff, C. (1989). The knowledge complexity of interactive proof systems. SIAM Journal on computing, 18(1), 186-208.
8. European Union. (2016). General Data Protection Regulation (GDPR)
9. Omotosho, O. (2024). The Relevance of Privacy-Preserving Techniques and Generative AI to DORA Legislation
10. Sharma, N. K. (2010). Data masking: A critical component of data security. International Journal of Computer Applications, 1(17), 42-47.
11. Costan, V., & Devadas, S. (2016, May). Intel SGX explained. In IACR Cryptology ePrint Archive (Vol. 2016, p. 86)
12. Woodcock, J., Larsen, P. G., Bicarregui, J., & Fitzgerald, J. (2009). Formal methods: Practice and experience. ACM Computing Surveys (CSUR), 41(4), 1-36.
13. Hoepman, J. H. (2007). Privacy enhancing technologies. In Privacy in Electronic Society (pp. 19-42). Springer, Boston, MA.
14. Thales. (2024) Data Threat Report, Global Edition

About the Authors

Dr Ilesh Dattani

The founder and CTO of Assentian. He has a first degree, a Masters in Mathematics and a PhD in Machine Learning. Over the last 20 years he has worked on the development of disruptive innovative new technologies in the financial services and insurance sectors. He sits on the International Standards Organisation committees on both Information Security and Artificial Intelligence. He was a founding member of the Whitechapel Think Tank – launched in 2016 by Barclays bank to accelerate innovation in financial services with emerging technologies (primarily Distributed Ledger Technologies). He has, and is, managing innovation activities in collaboration with global financial institutions and telecom providers to support their pilot projects using privacy preserving technologies where they are seeking to maximise data innovation whilst maintaining the highest levels of data privacy. He was an advisor to the European Central Bank during the investigation phase of the ongoing Digital Euro initiative. He is an advisor to the UK and Irish Governments on Innovation global alliances in the areas of Cyber Security, Fintech and AI. He is an investor in early-stage start-ups across Europe, Singapore and Australia and is a mentor to start-ups globally through his advisory work to accelerators IoT Tribe, CyberASAP, ICE71, CyRise and MACH-37.

Ollie Omotosho

Cloud, data and cybersecurity veteran of over 20 years. Having started from engineering degree, Manchester University, and transitioning to business management after postgraduate business qualifications, he has opened and grown global security technology markets for both blue chip and startup enterprises. His experience includes landmark projects such as digitisation of London transportation ticketing, national borders crypto security projects, critical national infrastructure security, bank peer-to-peer payments systems, securing smart meter rollouts, criminal justice system identity projects and blockchain-based industrial transformations. He is a volunteer advisor to Artificial Intelligence diversity NGO businesses, plus a contributor to the building of Quantum, AI and next generation finance initiatives globally, including build of labs, R&D and market-making joint propositions with the world's largest global service providers.