Ensuring Information Security and Data Governance in Cloud based Digital Contact Tracing Applications

Radhika Ravindranath
New Jersey, USA

ABSTRACT

The Covid-19 pandemic has stimulated the use of Digital Contact Tracing Applications(DCTAs) around the world, often implemented at a national scale. A public health crisis at such an unprecedented scale has accelerated research in the area of contact tracing, and the efficacy of digital contact tracing techniques. Given that there is now an estimated 2-3% chance of a pandemic striking at any given year, and a nearly 50% chance of a recurrence in the next 25 years it is important to learn from the lessons of past DCTAs. When cloud technologies are integrated when developing these applications, additional complexities related to cybersecurity, privacy and data governance arise.

This paper aims to identify and summarize the cybersecurity, privacy and ethical harms of cloud based centralized and decentralized DCTAs. The findings highlighted in this paper can help inform national and international security and privacy policies in the field of digital contact tracing, as well as allow organizations to embed security-by-design and privacy-by-design elements into their DCTA infrastructure.

Numerous national contact tracing systems were reviewed and their computing infrastructure, data collection, use and retention policies were studied in this paper. Potential cybersecurity, privacy and ethical harms associated with DCTAs were enumerated. International security and privacy standards like GDPR, NIST, ISO, etc. were reviewed to develop recommendations to address identified harms.

While DCTAs are essential for public health, they also come with significant risks to end user data security, privacy and ethical rights. Information security and privacy safeguards must be implemented in adherence to industry standards to minimize the risk. Especially when considering cloud based DCTAs, governments should prioritize platforms and libraries that offer strong technical features like encryption, access control, and compliance elements like regular auditing practices. Tailoring DCTAs to country or region specific needs is crucial. Carefully considering these factors will allow governing bodies to effectively utilize DCTAs while upholding end users' rights and maintaining public trust.

General Terms

Information Security, Data Privacy, Cybersecurity

Cloud Computing, Digital Contact Tracing, Data Governance, Ethical Considerations

Keywords

Cloud applications, digital contact tracing applications, privacy, cybersecurity, information security, data governance, ethical harms, data security

1. INTRODUCTION

The on-set of Covid-19 accelerated the deployment and utilization of DCTAs, with a majority of DCTAs being deployed at a national level. An estimated 321,332,010 people have downloaded one of 159 contact-tracing apps across 94 countries[1]. Contact tracing has applications beyond disease control, such as criminal investigations, emergency responses for disaster relief, marketing and research analytics and environmental studies like wildlife tracking and pollution monitoring. This study has utilized data from DCTAs adopted and used nationally during Covid-19.

There are currently numerous research publications on the efficacy of DCTAs[2,3] in the containment of diseases. However, more studies are required to assess the extent of personal or sensitive data collected through DCTAs, including cloud technologies, and the risks posed to individuals security, privacy, and ethical rights. This paper seeks to present research on the types of information systems which can be used for the development and deployment of DCTAs, the potential cybersecurity and privacy harms of the collection, storage and use of such systems, and defenses which can be employed to remediate such challenges. In this paper, cybersecurity refers to the protection of computer systems, networks and data from unauthorized access, in such a way that it affects the privacy and equity of the users of the computer system (DCTA). Privacy is the right to manage an individual's identity and how their data is collected, used, stored and shared in such a way that it adheres to a person's ethical rights.

2. APPROACH

As part of the discovery process, the following DCTAs from around the world and their associated data collection requirements were studied. It was noted that some countries used centralized computation systems models while others utilized distributed computer systems. As expected, the level of sensitive data collection also varied widely from country to country, with centralized systems tending to collect more personal information than decentralized systems. Details on use, storage and retention requirements for collected data varies regionally, as illustrated in the examples below.

Table 1. Digital contact tracing applications using cloud and on-premise deployment infrastructures

System Type	Centralized Infrastructure	Decentralized Infrastructure
Cloud- Based deployment	CoVID-19 Radar (India), Health Code App (China)	TraceTogether (Singapore), Corona-Warn-App (Germany), COVIDSafe (Australia)
On-prem deployment	COVID-19 Contact Tracing App (South Korea), TousAntiCovid (France)	N/A

The CoVID-19 Radar (India) collects a wider range of personal information, such as contact details, symptoms and location data. The National Institutes of Health published a paper in 2022 which focuses on how the app had several technical issues and inconsistencies in data privacy regulations, which deterred widespread adoption.

The Health Code App (China), for example is a centralized cloud based application which, developed by internet Giants Tencent and Alibaba utilized their existing user base and social network through Alipay or WeChat. Once registered, the health code app automatically pulled self reported travel histories and systems, pulled from travel and medical databases to determine if a user can use public spaces had potentially come into contact with an affected person, or needed to isolate.The app is tightly controlled by the Chinese government, raising surveillance and censorship concerns.

Decentralized Applications like Singapore's TraceTogther and Germany's Corona warn rely on bluetooth technology in order to log anonymous encounters. The decentralized nature of these applications balances data security and privacy by design with utility. Such applications may have a lower penetration depending on smartphone adoption and user behavior. It was further inferred that although cloud and on-prem infrastructure deployments may exhibit disparities in terms of development velocity, scalability, and disaster recovery protocols, the extent of application usage subsequent to download seems comparable across both models.

The approach of choosing the right technology, whether Bluetooth or GPS, can impact the accuracy and effectiveness of contact tracing. User trust is paramount in the success of contact tracing applications. Transparency and strong protections on how the collected data will be used, stored, processed and retained, as well as education on the clear benefits to public health are crucial in gaining public support.

2.1 Cloud vs On-prem solutions

Several factors must be considered when deciding on a cloud vs on-prem solution for DCTAs[4]. Since DCTAs are mainly used in disaster recovery situations or during a national or international health crises, advantages of cloud based DCTAs include scalability, cost effectiveness, rapid deployment and lighter maintenance updates as they will be handled by the cloud provider. Additionally, cloud based contact tracing systems offer security and privacy preserving APIs and libraries that adhere to standardized international frameworks. Some examples of frameworks include but are not limited to NIST Cybersecurity Framework, ISO 27001, Cloud Security Alliance (CSA) STAR registry, FedRAMP, GDPR (General Data Protection Regulation), Center for Internet Security (CIS) Controls, MITRE ATT&CK, HIPAA (Health Insurance Portability and Accountability Act), and PCI DSS (Payment Card Industry Data Security Standard. Privacy preserving APIs available include Federated Learning APIs (like TensorFlow Federated), Homomorphic encryption APIs (like Microsoft SEAL) and Differential Privacy APIs like Google's Differential Privacy Libraries. Cloud based deployments could also provide more readily available redundancies that can help protect user data in the case of regional natural disasters. However, cloud based contact tracing systems may have several disadvantages, such as dependency on the cloud provider, data privacy and sovereignty concerns and network latency impacts. It is important for the DCTA deployer to carefully evaluate the advantages and disadvantages of each approach to determine the best fit for their needs[5].

2.2 Centralized vs Decentralized systems

Centralized systems can be efficient for data collection and analysis, are scalable and offer easier government control over data. However decentralized systems can offer enhanced cybersecurity and data privacy while minimizing the collection and storage of personally identifiable information (PII) on central servers. Conversely, decentralized systems using bluetooth technologies have limitations in the granularity and quality of data collected and analyzed as compared to centralized digital contact tracing systems.

Centralized systems can be used to collect and infer more complete data, but raise privacy concerns and deter adoption. Decentralized systems, while less effective in some cases, prioritize user privacy and can lead to higher adoption rates[51]. Additionally, the choice of technology employed by the deployer (such as Bluetooth or GPS) can impact the adoption, effectiveness and accuracy of contact tracing applications.

Ultimately, the choice between centralized and decentralized systems depends on factors like regional needs, government trust, cultural definitions of privacy, and technological infrastructure. A hybrid approach combining elements of both systems may be effective in many cases as well.

Table 2. Key findings from comparative study of centralized and non-centralized digital contact tracing applications in cloud environments

System Type	**Centralized**	**Decentralized**
Cloud-Based	**Centralized Cloud**	**Decentralized Cloud**
On-Premises	Centralized On-Premises	Decentralized On-Premises
Advantages	Efficient data collection and analysis, scalability, centralized control	Enhanced privacy, resilience to single points of failure, reduced reliance on government
Disadvantages	Increased risk of data breaches and government surveillance, single point of failure	Potential for data fragmentation and interoperability issues, less efficient data analysis

3. CYBERSECURITY, PRIVACY AND ETHICAL HARMS DUE TO INFORMATION COLLECTED BY CONTACT TRACING APPLICATIONS

While contact tracing can be a valuable public tool, uncontrolled collection, storing and processing of personal data, beyond the purposes of the application itself, raises significant security, privacy and ethical concerns.

3.1 Privacy and Surveillance

Personal data collection, such as location data and travel history and contact information raises concerns of misuse for surveillance purposes. Large scale data collection to track individuals, such as in the case of Covid-19 can lead to mass surveillance efforts.[6,7,8]

3.2 Cybersecurity risk

Centralized data stores of personally identifiable information, or insecure use of data during collection, storage and use can pose cybersecurity risks, such as data breaches or malicious use[9,10].

3.3 Discrimination and Stigma

Individuals who test positive for a disease may be subjected to social stigma, targeted surveillance of already marginalized communities and leading to negative mental and physical health consequences[11]

3.4 Consent and Autonomy

Individuals must have the autonomy on whether or not to participate in contact tracing. Requiring individuals to participate in contact tracing programs can raise concerns about coercion and individual rights[12,13]

3.5 Access and equity

Digital contact tracing may not be available to marginalized or underserved communities, further alienating them from aid and exacerbating existing disparities

It is important that the above ethical concerns are taken into consideration when information security and privacy guardrails are implemented while building such systems. It is also important to involve stakeholders from diverse backgrounds in the development of the functional requirements for DCTAs.

4. IMPORTANT CONSIDERATIONS FOR MAINTAINING AN INFORMATION SECURITY AND PRIVACY PRESERVING APPROACH IN CONTACT TRACING

4.1 Data Collection

Before collecting any data, the DCTA must obtain explicit informed consent from users, explaining the purpose of the application, types of data collected, and how it will be used. This is in accordance with Article 4(11) of the GDPR which defines "consent" as "any freely given, specific, informed and unambiguous indication of the data subject's will, by which the data subject agrees, either by a statement or by a clear affirmative action, to the processing of personal data relating to him or her[14].

Only the necessary amount of data must be collected, such as location data, Bluetooth device encounters and basic demographic information to minimize identifiability. The collection, storage, use and processing of this data must not be used in excess of what the user has consented to[15].

Preserve anonymity or pseudonymity of the users in the application to reduce stigma and equity[16].

4.2 Processing

Secure Processing methodologies, like cryptographic algorithms, homomorphic encryption[17] and secure multi-party computation[18] must be utilized to prevent data breaches . Additionally use other security measures like 2FA, Access control lists and intrusion detection systems at the processing site in the case of centralized data processing facilities[19,20]

Minimize data joining with other personally identifiable information, which can increase identifiability of the individuals.. Replace PII with anonymous or pseudonymous identifiers to protect privacy. Rely on aggregated data results instead of retaining personal information.

4.3 Use

Collected data must strictly be used only for the purpose of contact tracing. Regular backups must be implemented to protect against data corruption or loss. Users must be notified of any changes in data use in order to maintain data use transparency[21,22,23,24].

4.4 Storage and Data Retention

Establish clear data retention policies[22] and set up automated data deletion controls[25] to promptly delete data from all information systems when no longer needed

Implement regular back ups[26,27] to in the case of disaster recovery needs and to minimize any accidental or malicious data loss or corruption

Minimize the amount of data stored and retain only what is required for contact tracing

4.5 Notice and Control

Provide clear and easy to understand privacy notices for users so they can understand what data is collected, and how it is processed, used and retained[28].

Ensure that any third parties involved go through vetted Vendor selection processes to ensure that the standard of security and privacy is maintained, Minimize the sharing of any sensitive information with third parties[29,30].

4.6 Access Control and Transparency

Restrict access[31,32] to raw data and rely on the use of service accounts for any processing. Access to the aggregated, process data must also be minimized and audited. Role based access control[33,34] (RBAC) may be employed in centralized contact tracing systems, while permissioned block chains and DDAC frameworks can be employed in decentralized systems. Cloud based systems have built in Identity and Access management (IAM) systems, for example, AWS IAM, Azure Active Directory, Google Cloud IAM

The use of strong authentication systems[35,36], such as multi factor authentication must be used to protect access to the data

4.7 Additional considerations

Cross border data transfers must take into account local privacy policies or lack thereof[37]

Incident response plans must be developed in the case of disaster recovery and data breaches, to protect data and promptly notify affected individuals[38,39]

5. ADDITIONAL INFORMATION SECURITY TECHNIQUES FOR PRESERVING SECURITY AND PRIVACY IN CONTACT TRACING

Anomaly detection through ML/AI which can be used to identify patterns or anomalies which can indicate breaches or unauthorized access. They can also be used to anonymize or pseudonymous the data for analysis by identifying patterns.

Federated Analytics[40], in which data is kept and processed on individual devices can reduce the risk of breaches. The ML models are trained locally on devices, which minimizes sensitive data exposure and increases the efficiency of processing, especially for large- scale, centralized approaches

Differential privacy[41], which includes adding noise to the data, making it more difficult to identify individual data points. This approach can still pose privacy risk, but the level of noise can help balance the level of privacy protection with the utility of the data, From a security standpoint as well, it can make it more challenging for attackers to glean valuable information from the data

Data Encryption[42,43] at rest and in transit - Cloud Providers offer robust encryption mechanisms to protect data at rest. Some examples of Key Management facilities by cloud providers are Amazon Web services (AWS) KMS, Microsoft Azure Key Vault, Google Cloud KMS[44]

Homomorphic encryption, with which computations can be performed on encrypted data without the additional step of decryption, can be computationally expensive, but preserves the confidentiality, integrity and availability of the data from malicious insiders and outsiders.

Data Isolation in the cloud can allow DCTA deployers to create isolated virtual networks within the cloud to segregate data regionally and improve security. Examples of cloud solutions include AWS VPC, Azure Virtual Networks, Google Cloud VPC.

Cloud providers also comply with industry standards such as ISO 27001, HIPAA and GDPR, and can offer solutions that allow organizations to manage compliance. AWS Compliance center Azure Trust Center and Google Cloud Trust center are some examples of cloud compliance management solutions.

6. CONCLUSIONS

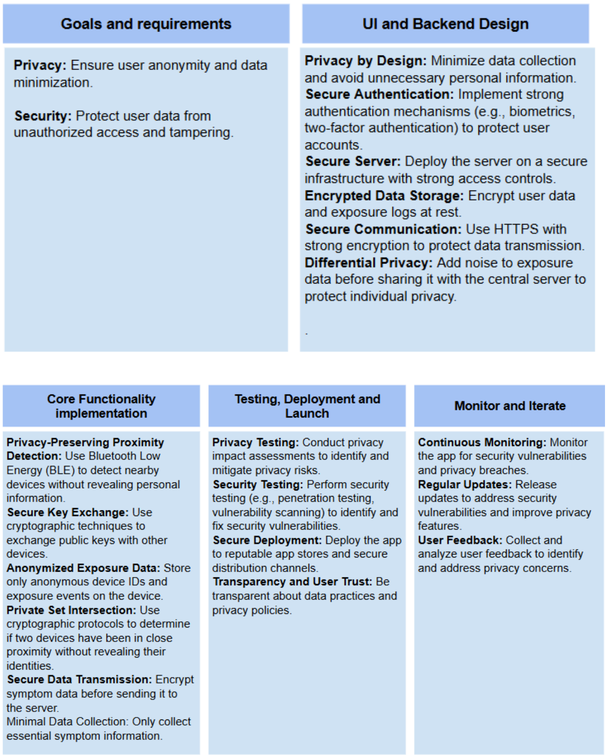
Digital contract tracing, while valuable in public health and disaster recovery, present significant challenges to user data security, privacy and ethical rights. The widespread use of these applications during the COVID-19 pandemic underscores the need for careful collection of data, and the potential risks associated with its processing, storage and use. Necessary information security and privacy safeguards must be implemented to mitigate these risks.

This paper explores the technical infrastructure of the information systems used in contact tracing deployments, such as cloud, on-prem, centralized and decentralized, with a focus on cloud based systems. While each approach offers unique pros and cons, careful considerations must be made to tailor DCTAs based on the specific needs of the country they are deployed in.

Addressing concerns like surveillance risk, cyber security risks, privacy risks and ethical concerns like stigma and equity are important in ensuring that DCTAs remain beneficial while protecting individual rights.

Industry standards must be used to implement robust privacy and information security measures to mitigate the identified risks.This includes data minimization strategies, informed consent, secure data processing, data sharing limitations, retention policies, access controls, and anonymization, aggregation and pseudonymization techniques. Advanced techniques like anomaly detection using Machine Learning and Artificial intelligence, federated analytics, differential privacy, cryptographic techniques and policy compliance further enhances security and privacy. Specifically, cloud technologies have inbuilt libraries which can assist in easy, standardized deployment of such security and privacy controls.

One way of ensuring that privacy and safety goals are met can be by dividing the application development process into distinct phases and mapping out specific privacy and security goals, deployers can ensure that security by design and privacy by design requirements are embedded into DCTAs. An example methodology is illustrated below.



Goals and requirements

Privacy: Ensure user anonymity and data minimization.

Security: Protect user data from unauthorized access and tampering.

UI and Backend Design

Privacy by Design: Minimize data collection and avoid unnecessary personal information.
Secure Authentication: Implement strong authentication mechanisms (e.g., biometrics, two-factor authentication) to protect user accounts.
Secure Server: Deploy the server on a secure infrastructure with strong access controls.
Encrypted Data Storage: Encrypt user data and exposure logs at rest.
Secure Communication: Use HTTPS with strong encryption to protect data transmission.
Differential Privacy: Add noise to exposure data before sharing it with the central server to protect individual privacy.

.

Core Functionality implementation

Privacy-Preserving Proximity Detection: Use Bluetooth Low Energy (BLE) to detect nearby devices without revealing personal information.
Secure Key Exchange: Use cryptographic techniques to exchange public keys with other devices.
Anonymized Exposure Data: Store only anonymous device IDs and exposure events on the device.
Private Set Intersection: Use cryptographic protocols to determine if two devices have been in close proximity without revealing their identities.
Secure Data Transmission: Encrypt symptom data before sending it to the server.
Minimal Data Collection: Only collect essential symptom information.

Testing, Deployment and Launch

Privacy Testing: Conduct privacy impact assessments to identify and mitigate privacy risks.
Security Testing: Perform security testing (e.g., penetration testing, vulnerability scanning) to identify and fix security vulnerabilities.
Secure Deployment: Deploy the app to reputable app stores and secure distribution channels.
Transparency and User Trust: Be transparent about data practices and privacy policies.

Monitor and Iterate

Continuous Monitoring: Monitor the app for security vulnerabilities and privacy breaches.
Regular Updates: Release updates to address security vulnerabilities and improve privacy features.
User Feedback: Collect and analyze user feedback to identify and address privacy concerns.

In the above example, the development lifecycle has been broken down into 5 phases namely, Goals and Requirements; UI and backend design; Core functionality implementation; Testing deployment and launch; Monitor and Iterate. Each phase has distinct privacy and security goals identified, and are embedded into the technical design. Carefully considering these factors will allow governing bodies to effectively utilize DCTAs while maintaining public trust.

7. REFERENCES

[1] Comparitech. (2024, October 17). Contact-tracing app adoption by country. [Online]. Retrieved October 17, 2024, from https://www.comparitech.com/blog/vpn-privacy/contact-tracing-app-adoption-by-country/

[2] A. Wesolowski, C. A. Eagle, J. J. O. Jr., N. M. Smith, and J. S. Salathé, "Nationwide rollout reveals efficacy of epidemic control through digital contact tracing," Nat. Commun., vol. 12, no. 1, p. 5918, Dec. 2021, doi: 10.1038/s41467-021-26144-8.

[3] G. Cencetti, G. Santin, A. Longa, E. Pigani, A. Barrat, C. Cattuto, et al., "Digital proximity tracing on empirical contact networks for pandemic control," Nat. Commun., vol. 12, no. 1, p. 1655, Dec. 2021, doi: 10.1038/s41467-021-21809-w.

[4] National Institute of Standards and Technology (NIST). (2011, September). NIST Cloud Computing Reference Architecture (Special Publication 500-292). [Online]. Available: https://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication500-292.pdf

[5] Microsoft. (2016, August). Windows Azure Security, Privacy, and Compliance. [Online]. Available: https://download.microsoft.com/download/1/6/0/160216AA-8445-480B-B60F-5C8EC8067FCA/WindowsAzure-SecurityPrivacyCompliance.pdf

[6] Mayer, S. E., & Marwell, G. (2005). Surveillance and privacy: A paradoxical relationship. Social Science Quarterly, 86(3), 723-744.

[7] Solove, D. (2001). Surveillance: A broader understanding. Stanford Law Review, 54(3), 579-643.

[8] Turow, S. (2009). The privacy paradox: Social tracking in the digital age. Yale Law & Policy Review, 27(2), 347-403.

[9] Russell, D., & Schneier, B. (2011). Security Engineering: A Guide to Building Trustworthy Systems. Addison-Wesley Professional.

[10] Stallings, W., & Brown, L. (2017). Computer Security: Principles and Practice. Pearson Education.

[11] B. G., & Phelan, J. C. (2001). Stigma and the social construction of illness. Journal of Health and Social Behavior

[12] European Union Agency for Cybersecurity (ENISA). (2020). Contact Tracing and Data Protection: A Guide for Policymakers and Practitioners. ENISA Report.

[13] European Union Agency for Cybersecurity (ENISA). (2020). Contact Tracing and Data Protection: A Guide for Policymakers and Practitioners. ENISA Report.

[14] GDPR, Art. 4(11). "Consent" means any freely given, specific, informed and unambiguous indication of the data subject's will, by which the data subject agrees, either by a statement or by a clear affirmative action, to the processing of personal data relating to him or her.

[15] GDPR Recital 39: Personal data should be adequate, relevant and not excessive in relation to the purposes for which they are processed.

[16] GDPR Recital 26: Pseudonymization should be considered where technically feasible and appropriate in order to reduce the risk to data subjects.

[17] Gentry, C. (2009). Fully homomorphic encryption using ideal lattices. In Proceedings of the 41st Annual ACM Symposium on Theory of Computing (STOC) (pp. 169-178). ACM

[18] Goldreich, O., Micali, S., & Wigderson, A. (1987). How to play any mental game. In Proceedings of the 19th Annual ACM Symposium on Theory of Computing (STOC) (pp. 218-229). ACM.

[19] NIST. (2015). NIST Special Publication 800-171: Revised: Protecting Controlled Unclassified Information in Non-Federal Information Systems and Organizations. National Institute of Standards and Technology

[20] ISO. (2013). ISO/IEC 27001:2013 Information security management systems -- Requirements for application to information security management systems. International Organization for Standardization.

[21] GDPR, Art. 5(1)(b): Personal data shall be processed fairly and lawfully.

[22] GDPR, Art. 5(1)(e): Personal data shall be kept in a form which permits identification of data subjects for no longer than is necessary for the purposes for which the personal data are processed.

[23] NIST Special Publication 800-171: NIST. (2015). NIST Special Publication 800-171: Revised: Protecting Controlled Unclassified Information in Non-Federal Information Systems and Organizations. National Institute of Standards and Technology.

[24] ISO 27001: ISO. (2013). ISO/IEC 27001:2013 Information security management systems -- Requirements for application to information security management systems. International Organization for Standardization.

[25] NIST SP 800-171, Control 3.8: Data retention and disposition procedures.

[26] NIST SP 800-171, Control 3.7: Backup procedures.

[27] ISO 27001, A.16.1.1: Backup and recovery procedures.

[28] GDPR Art. 13(1): When collecting personal data from the data subject, the controller shall provide the data subject with certain information at the time of collection, including the identity and contact details of the controller, the purposes of the processing, the categories of personal data concerned, the intended recipients of the personal data, the envisaged period of storage of the personal data, the existence of a right of access and rectification or erasure, the existence of a right to complain to a supervisory authority, and the source of the personal data if it is not collected from the data subject.

[29] GDPR Art. 28: The controller shall, by contract or other legally binding instrument, ensure that any processor it engages is subject to the same obligations as the controller under the GDPR.

[30] NIST SP 800-171, Control 3.3: Third-party service provider agreements.

[31] NIST SP 800-171, Control 3.4: Access control procedures.

[32] ISO 27001, A.12.4.1: Access control policies.

[33] NIST SP 800-171, Control 3.5: Role-based access control.

[34] ISO 27001, A.12.5.1: Role-based access control.

[35] NIST SP 800-171, Control 3.6: Authentication procedures.

[36] ISO 27001, A.12.3.1: Authentication procedures.

[37] GDPR Art. 44: Personal data may only be transferred to a third country or an international organization if adequate safeguards are in place, such as standard contractual clauses adopted by the Commission.

[38] GDPR Art. 33: When a personal data breach is likely to result in a high risk to the rights and freedoms of data subjects, the controller shall communicate the breach to the relevant supervisory authority without undue delay.

[39] NIST SP 800-171, Control 3.9: Incident response procedures

[40] Kairouz, P., McMahan, H. B., Avent, B., et al. (2019). Advances in federated learning. arXiv preprint arXiv:1906.00582.

[41] Dwork, C., McSherry, F., Nissim, K., & Smith, A. (2006). Calibrating noise to privacy. In Theory of Cryptography Conference (TCC) (pp. 265-284). Springer.

[42] NIST Special Publication 800-171, Control 3.7: Backup procedures.

[43] ISO 27001, A.13.2.1: Physical security measures.

[44] The Pandemics to Come. Boston College Magazine, Winter 2022. (Author and DOI unavailable)