Deloitte. 德勤

Data sharing in financial services:

Five techniques to enhance privacy and confidentiality

Introduction

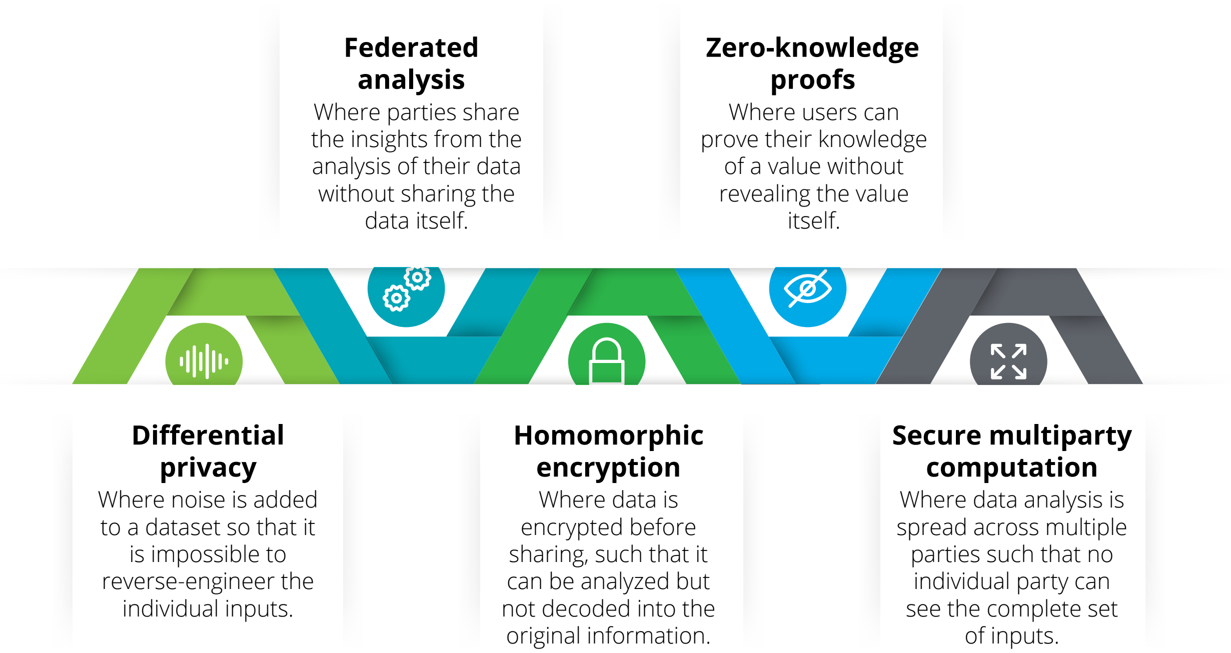
In financial services, data sharing is fraught with tension. On the one hand, it can help fight transaction fraud, deliver more personalized advice to customers, and detect the buildup of systemic risks. On the other hand, customers are increasingly wary about how their data is stored and used—and, as reforms like the EU's General Data Protection Regulation and the UK's Open Banking show, regulators are inclined to agree.

That, in a nutshell, highlights the competing obligations surrounding privacy: there's value in sharing data, but protecting privacy and confidentiality is a critical responsibility of any financial institution.

Since 2015, Deloitte has worked with The World Economic Forum to gauge the forces of change in financial services. In the most recent phase—which will be reported in the forthcoming report *Navigating uncharted waters: A roadmap to responsible innovation with AI in financial services*—we discovered these competing obligations surrounding privacy and data sharing. This in turn led to a deeper examination of ways to unlock the value that shared data can provide without threatening privacy and confidentiality.

Privacy enhancing techniques

This report explores five key "privacy enhancing techniques":



Federated analysis
Where parties share the insights from the analysis of their data without sharing the data itself.

Zero-knowledge proofs
Where users can prove their knowledge of a value without revealing the value itself.

Differential privacy
Where noise is added to a dataset so that it is impossible to reverse-engineer the individual inputs.

Homomorphic encryption
Where data is encrypted before sharing, such that it can be analyzed but not decoded into the original information.

Secure multiparty computation
Where data analysis is spread across multiple parties such that no individual party can see the complete set of inputs.

The report also provides a high-level overview of how each technique works, the types of data sharing problems they can be used to solve, and the subsectors of financial services in which they are most immediately applicable.

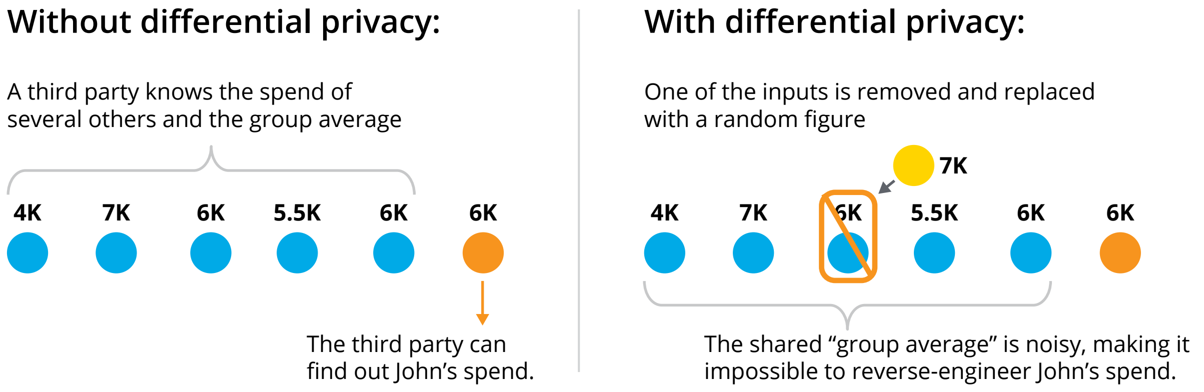
Differential privacy

A common belief is that anonymizing personally identifiable information (PII) is enough to protect customers' privacy, but this isn't always the case.

To understand why, suppose John Doe shares his bank account data with a personal financial advisory app. This app makes it easier for customers to manage their spending and compare it with similar customers. John asks the app to compare what he spends in bars annually with the average for his demographic. The app returns an aggregate response: "Males aged 25-29 in this zip code generally spend $5,750 a year in bars."

However, suppose a bad actor wanted to find out how much John is spending in bars. The bad actor could accomplish this by, for example, changing their own address to fit within John's demographic. By then querying the system again knowing some of the inputs (i.e., their own) and cross-referencing with other data (e.g., census data), this third party could breach John's privacy and deduce his bar spend.

To prevent this kind of breach, the system can add noise to its calculation of the average, using differential privacy to measure how much noise is necessary to achieve the desired level of privacy. For instance, it could replace one customer's spend with a random number, changing the reported average enough to make it impossible to reverse-engineer the inputs while producing a useful statistic for honest users.



Without differential privacy:

A third party knows the spend of several others and the group average

The third party can find out John's spend.

With differential privacy:

One of the inputs is removed and replaced with a random figure

The shared "group average" is noisy, making it impossible to reverse-engineer John's spend.

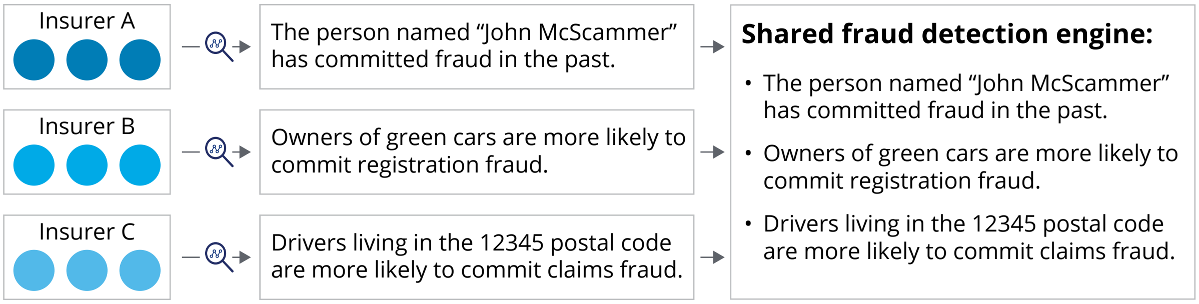
Differential privacy holds particular promise for retail banks, insurers, payment service providers and other institutions that maintain sensitive personal data. The technique can enable these institutions to aggregate and analyze sensitive data without risking the privacy of the customers they serve.

Federated analysis

Sometimes, the data needed to make a decision is scattered across multiple sources (e.g., identifying fraud networks spread across multiple banks). It can be more efficient to combine the data into a single database for easier analysis, but this may not always be possible. If the data is internal but split across jurisdictions, for instance, privacy restrictions may prevent its transfer. And if the data is shared across institutions, customers may object to releasing their private information and institutions may worry about how third parties would handle the data, particularly if they happen to be competitors.

One way to address these issues is to analyze each dataset separately and build several independent models, then combine these intermediate decisioning models into a single aggregated system—a technique known as federated analysis. For example, consider several insurance companies seeking to detect fraud across their systems. They can independently analyze their data, then share only their insights with each other. This allows them to benefit from one another's learnings without threatening the privacy of their customers.

With federated analysis:



The person named "John McScammer" has committed fraud in the past.

Owners of green cars are more likely to commit registration fraud.

Drivers living in the 12345 postal code are more likely to commit claims fraud.

Shared fraud detection engine:

- The person named "John McScammer" has committed fraud in the past.
- Owners of green cars are more likely to commit registration fraud.
- Drivers living in the 12345 postal code are more likely to commit claims fraud.

This technique is already embedded into other organizations' analytical systems. For example, large technology companies use federated analysis (and other privacy enhancing techniques) to power the "next word" recommendations built into the keyboards on their mobile phone operating systems.

Federated analysis is a way for financial institutions to break down key barriers to getting insights from multiple private datasets. For instance, federated analysis could encourage greater use of connected devices that promote responsible behavior among insurance customers (think auto and fitness trackers), in part by assuring those customers that their sensitive data never leaves their phones. Meanwhile, insurers could still capture the aggregate insights from their customers' data. In sectors like payments and insurance, federated analysis can also boost security by letting rival institutions participate in a common fraud detection network that doesn't expose their internal data.

Homomorphic encryption

Sometimes a financial institution—or one of its customers—would like to engage a third party for data analysis. The third party might have complementary data or proprietary analytics the institution doesn't have. However, the data steward or owner may lack permission to transfer the data or have concerns about keeping the data safe.

Homomorphic encryption (HE) can bridge this gap by encrypting data so that it can be analyzed without knowing the underlying information. With HE, it isn't necessary to decrypt the data first. Neither can anyone other than the intended party read the results of the analysis.

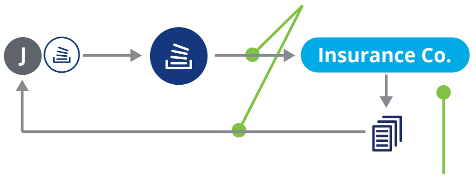
Consider a situation where John Doe would like to see if his medical history reveals any potential health risks. His health insurance provider has a technology services unit with the capabilities to run such an analysis, but John Doe wants to maintain the confidentiality of his health records.

With HE, John Doe can encrypt the data and send it to his insurer while holding on to the key. The technology unit can run the data through its models without having to know what is in the records or the results, then return both to John Doe to unlock and read.

Without homomorphic encryption:

John places his health records in a box, ships them to the company, which analyzes them to produce a report and ships it back to John.

Data could be maliciously accessed in transportation.

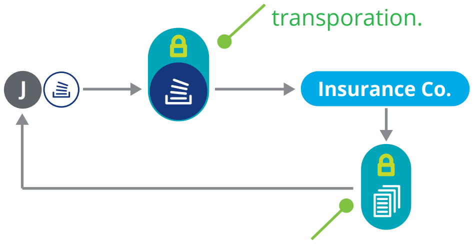


Data could be maliciously accessed by the company itself or an external bad actor who gains access to the office.

With homomorphic encryption:

John's health records are homomorphically encrypted prior to sharing, making it difficult for anyone but him to see the data or the results of any subsequent analysis.

Data is secure during transporation.



The company conducts its analysis without being able to see the underlying data at any point.

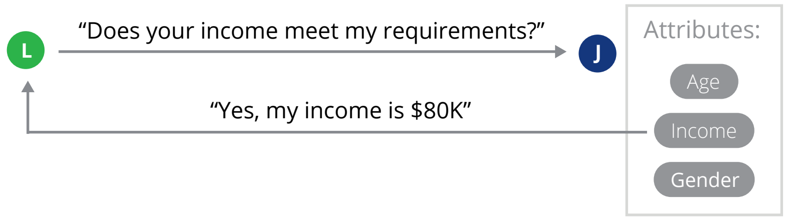
HE is potentially useful to any financial institution interested in analyzing sensitive data on the cloud or via third-party capabilities. Today, these options are limited due to concerns about data breaches, localization requirements, and privacy regulation. But that could change with HE solutions that provide a practical way to keep data encrypted and safe from prying eyes, even while it's in use.

Zero-knowledge proofs

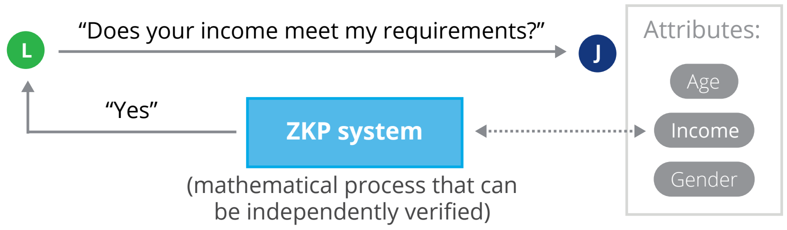
Many customers would rather not reveal more than is absolutely necessary to complete a transaction, lest the information be used against them. For instance, let's say John must show a landlord he can afford to rent an apartment. But John doesn't want the landlord to know that he makes a lot more than the required minimum and risk the landlord raising the rent at the first available opportunity.

John's bank can help by using a technique called zero-knowledge proof (ZKP). With ZKP, the bank uses a mathematical proof to verify to the landlord that John earns enough to afford the rent, without revealing his actual income. Because it's automated, John can qualify himself quickly, without getting bank personnel involved.

Without zero-knowledge proofs:



With zero-knowledge proofs:

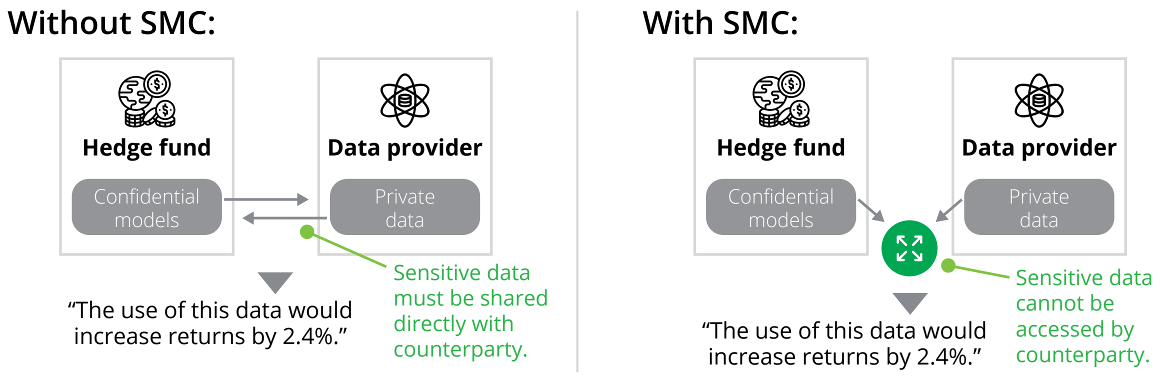

L
"Does your income meet my requirements?"
J
"Yes"
ZKP system
(mathematical process that can be independently verified)
Attributes:
Age
Income
Gender

Institutions large and small are increasingly using ZKP in payments, infrastructure, self-sovereign digital identity solutions, and more. This use is driving a broader shift toward "zero-knowledge architectures," where institutions design their data systems to be able to access only the minimum information necessary for their given tasks and maintain the privacy of all other data.

Secure multiparty computation

Secure multiparty computation (SMC) allows institutions to jointly analyze data without any one institution being able to access the complete dataset. This allows multiple institutions with sensitive information to work together to create value without risking their confidential information.

Consider the following example: A hedge fund seeks to purchase data from a third-party data provider to improve the quality of its trading models. The hedge fund wants to know that the data would actually be helpful before making the purchase. At the same time, the third party is hesitant to share their data before payment. Traditionally, the two firms would share a historical dataset (which may not be representative of the present-day performance) or a small sample set (which may be difficult to integrate into the hedge fund's models and accurately represent the value of the data).


Without SMC:
Hedge fund
Confidential models
Data provider
Private data
"The use of this data would increase returns by 2.4%."
Sensitive data must be shared directly with counterparty.
With SMC:
Hedge fund
Confidential models
Data provider
Private data
"The use of this data would increase returns by 2.4%."
Sensitive data cannot be accessed by counterparty.

SMC can be used to combine these two sensitive aspects—the hedge fund's models and the provider's data—and compute the value of the data, without either party being able to access the other's confidential information. This way, the hedge fund can make a more informed decision about whether to buy the data without the two parties having to trust each other. Meanwhile, each party can independently audit the SMC system to ensure it's protecting the privacy of the input data.

In short, SMC is an enabling technique for situations where multiple institutions each hold part of the answer to a common problem, but none of them wants others to access their own data. One sector where this is notably relevant is capital markets, due to the amount of proprietary data that can inform trading and investment. And like federated analysis, SMC can enable the development of fraud detection networks across institutions.

Taking privacy to the next level

Financial institutions have a long history of weighing the utility of data sharing with the obligation to maintain privacy and confidentiality. Now, five relatively nascent technologies have the potential to fundamentally alter these dynamics.

What they have in common is an ability to allow institutions, customers, and regulators to analyze data and distribute the resulting insights without having to share the underlying data itself. This way, they can greatly reduce the risks associated with data sharing. The result? New ways for financial institutions to address their biggest, most pressing problems in a way that is acceptable to customers, regulators, and societies at large.

This article is derived from *The Next Generation of Data-Sharing in Financial Services: Using Privacy Enhancing Techniques to Unlock New Value*, prepared by the World Economic Forum in collaboration with Deloitte. The World Economic Forum will continue to explore the effects of change in financial services. If you'd like to discuss the ideas in this report—formally or informally—we'd like to hear from you.

Global Contacts

Bob Contri
Financial Services Industry Leader, Deloitte Global
bcontri@deloitte.com

Rob Galaski
Global Leader, Banking & Capital Markets, Deloitte Consulting
rgalaski@deloitte.ca

Authors

Ishani Majumdar
Senior Consultant, Omnia AI, Deloitte Canada
ismajumdar@deloitte.ca

Hemanth Soni
Senior Consultant, Monitor Deloitte, Deloitte Canada
hemasoni@deloitte.ca

We would also like to thank **Courtney Kidd Chubb** and **Denizhan Uykur** from Deloitte Canada for their contributions to this report.

China Contacts

Fang Ye
FSI Risk Advisory Leader, Mainland China
+86 21 6141 1569
yefang@deloitte.com.cn

Tony Wood
FSI Risk Advisory Leader, Hong Kong
+852 2852 6602
tonywood@deloitte.com.hk

Tonny Xue
Partner, Risk Advisory
Cyber Security and Technology Risk Service Leader
+86 10 8520 7315
tonxue@deloitte.com.cn

Tommy He
Partner, Risk Advisory
Cyber Security and Technology Risk Service, FSI Co-Lead
+86 10 8512 5312
the@deloitte.com.cn

Eva Kwok
Partner, Risk Advisory
Cyber Security and Technology Risk Service, FSI Co-Lead
+852 2852 6304
evakwok@deloitte.com.hk

Frank Xiao
Partner, Risk Advisory
Cyber Security and Technology Risk Service,
Data Security and Privacy Leader
+86 10 8512 5858
frankxiao@deloitte.com.cn

Vivi He
Partner, Risk Advisory
Cyber Security and Technology Risk Service
+86 755 3353 8697
vhe@deloitte.com.cn

David Jiang
Partner, Risk Advisory
Cyber Security and Technology Risk Service
+86 21 2312 7088
davidjiang@deloitte.com.cn

关于德勤

Deloitte (“德勤”) 泛指一家或多家德勤有限公司，以及其全球成员所网络和它们的关联机构。德勤有限公司 (又称“德勤全球”) 及其每一家成员所和它们的关联机构均为具有独立法律地位的法律实体。德勤有限公司并不向客户提供服务。请参阅www.deloitte.com/cn/about了解更多信息。

德勤亚太有限公司 (即一家担保有限公司) 是德勤有限公司的成员所。德勤亚太有限公司的每一家成员及其关联机构均为具有独立法律地位的法律实体，在亚太地区超过100座城市提供专业服务，包括奥克兰、曼谷、北京、河内、香港、雅加达、吉隆坡、马尼拉、墨尔本、大阪、上海、新加坡、悉尼、台北和东京。

德勤于1917年在上海设立办事处，德勤品牌由此进入中国。如今，德勤中国为中国本地和在华的跨国及高增长企业客户提供全面的审计及鉴证、管理咨询、财务咨询、风险咨询和税务服务。德勤中国持续致力为中国会计准则、税务制度及专业人才培养作出重要贡献。德勤中国是一家本土注册成立的中国专业服务机构，由德勤中国的合伙人所拥有。敬请访问www2.deloitte.com/cn/zh/social-media，通过我们的社交媒体平台，了解德勤在中国市场成就不凡的更多信息。

本通信中所含内容乃一般性信息，任何德勤有限公司、其成员所或它们的关联机构 (统称为 “德勤网络”) 并不因此构成提供任何专业建议或服务。在作出任何可能影响您的财务或业务的决策或采取任何相关行动前，您应咨询合资格的专业顾问。任何德勤网络内的机构均不对任何方因使用本通信而导致的任何损失承担责任。

Designed by CoRe Creative Services. RITM0374219

这是环保纸印刷品