\Orchestrating a brighter world NEC

セキュアなデータ結合のための秘密計算技術と制度の方向性

NEC セキュリティ研究所
竹之内隆夫

2018年2月22日 PWS Meetup@明治大学

はじめに

本日の発表の趣旨

- 匿名加工情報では実現できない複数組織でのデータ結合を秘密計算技術で安全に実現できるのでは？

CSS/PWS2017@山形にて、秘密計算技術を用いたセキュアなデータ結合に関する企画セッションを実施

- 1. 新潟大学 須川先生（座長）： セッションの趣旨説明
- 2. NTT 藤村さん：ユースケース・法制度課題など
- 3. 日立コンサルティング 美馬さん：準同型暗号方式の秘密計算技術の説明
- 4. **NEC 竹之内**：秘密計算技術の整理
- 5. ひかり総合法律事務所 板倉先生：法制度の方向性について

CSS/PWS2017での企画セッション内容をまとめつつ、
技術説明や今後の展望を説明

目次

1.複数組織のデータ結合・分析への期待

組織間でのデータ利活用への期待

データを組織間で共有し、データを結合・分析することは、社会的な価値を創出できると期待されているが、プライバシ等の課題が存在

データの共有分析による価値創出

- 例：米国の医療は、病院・介護者・製薬会社などがデータを共有して活用できれば、**毎年3000億ドル以上**の価値を生み出せる。[1][2]

データ共有の狙い	創出価値(億ドル)
生活習慣の改善	700-1000
医療・介護の連携	900-1100
最適な医療の選定	500-700
費用対効果の検証	500-1000
創薬・実証の加速	400-700
合計	3000-4500

データ共有の主な阻害要因

1. 個人のプライバシー保護：
   個人の同意なく、個人情報の**第三者への提供は禁止**

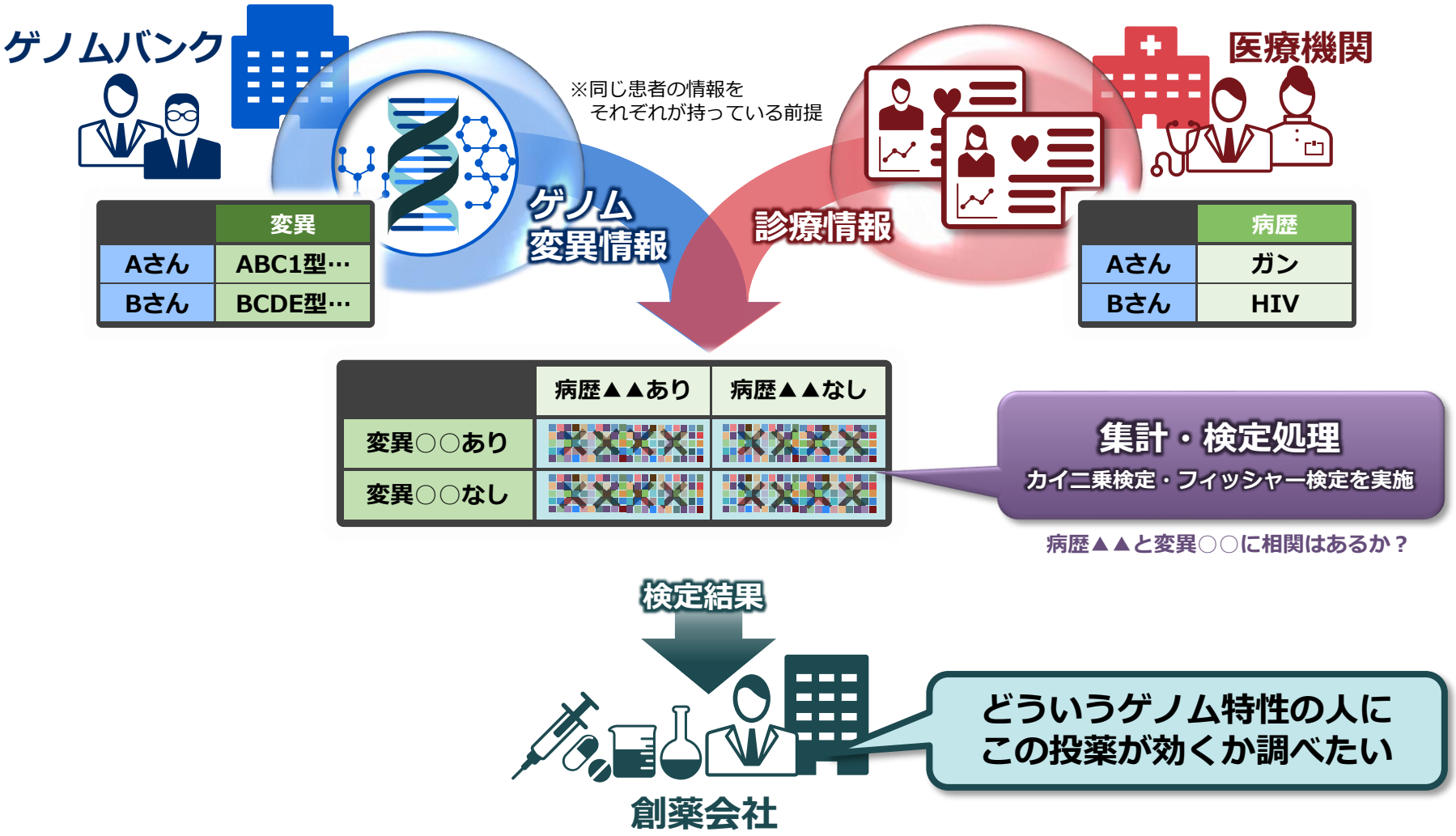
2. 競争力の源泉になる秘密情報：
   企業や研究機関は、データを**競合に開示するのを嫌う**

[1] McKinsey Global Institute, *Big data: The next frontier for innovation, competition, and productivity*, 2011年5月

[2] McKinsey Global Institute, *The 'big data' revolution in healthcare — Accelerating value and innovation*, 2013年1月. Exhibit 4.

複数組織のデータ結合分析の一例：医学分析

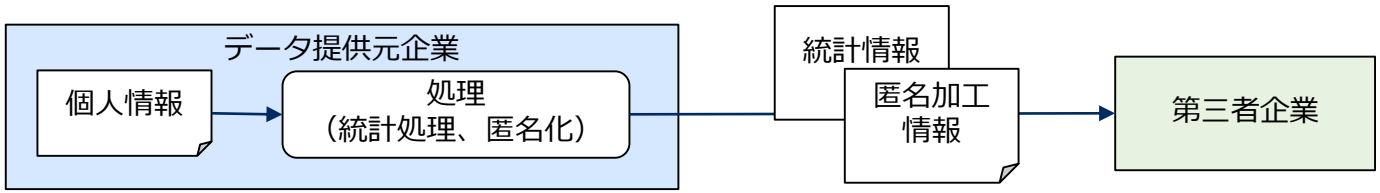
複数の医療機関の同一人物のデータを結合して分析し、医学研究に活用



匿名化技術では不十分

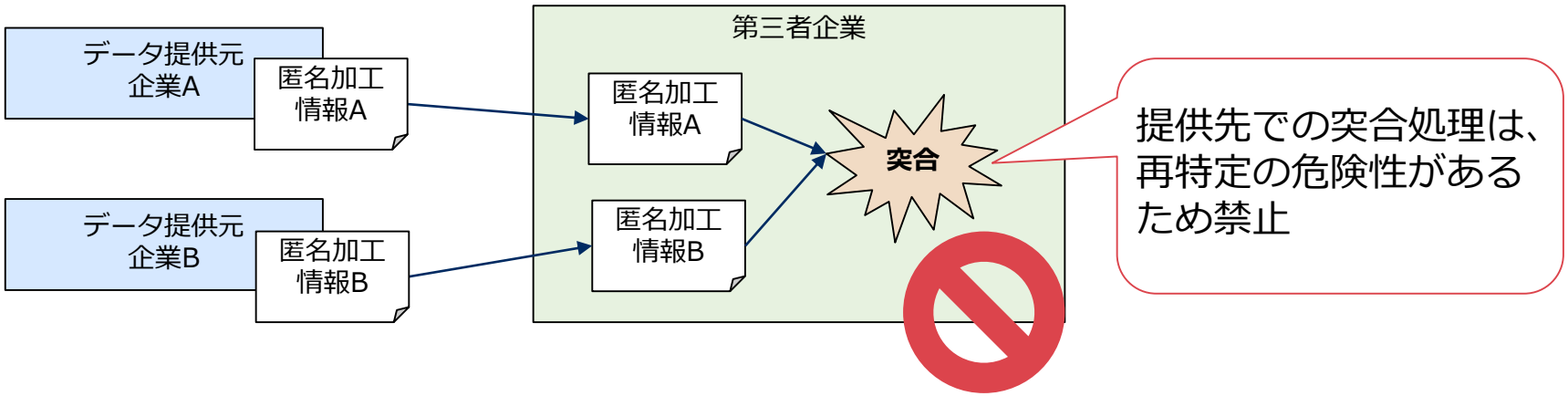
匿名化技術では、組織間でのデータ結合分析への対応は困難

匿名加工情報や統計情報であれば一定の条件下※1で第三者企業へ提供可能



※1: オプトアウトへの対応、提供先での突合処理の禁止、など

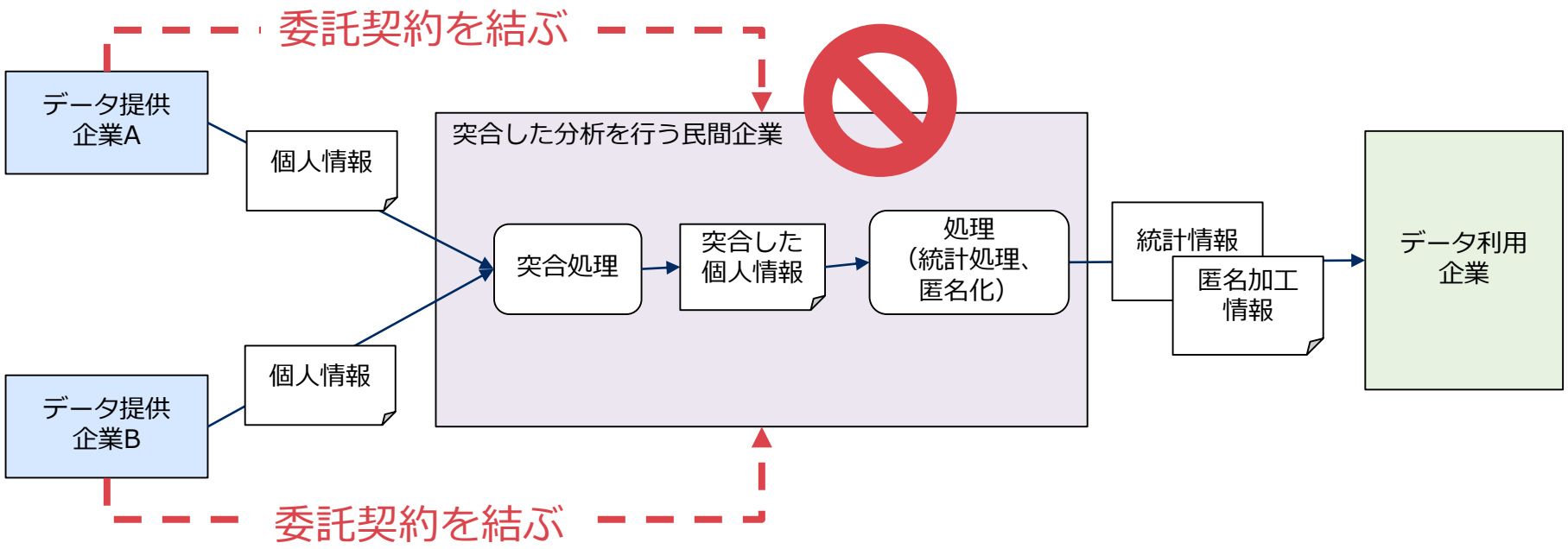
匿名化は、複数組織のデータを結合することは想定していない



委託契約でもデータ結合は禁止

委託契約先で異なる委託元の個人情報を突合することは禁止

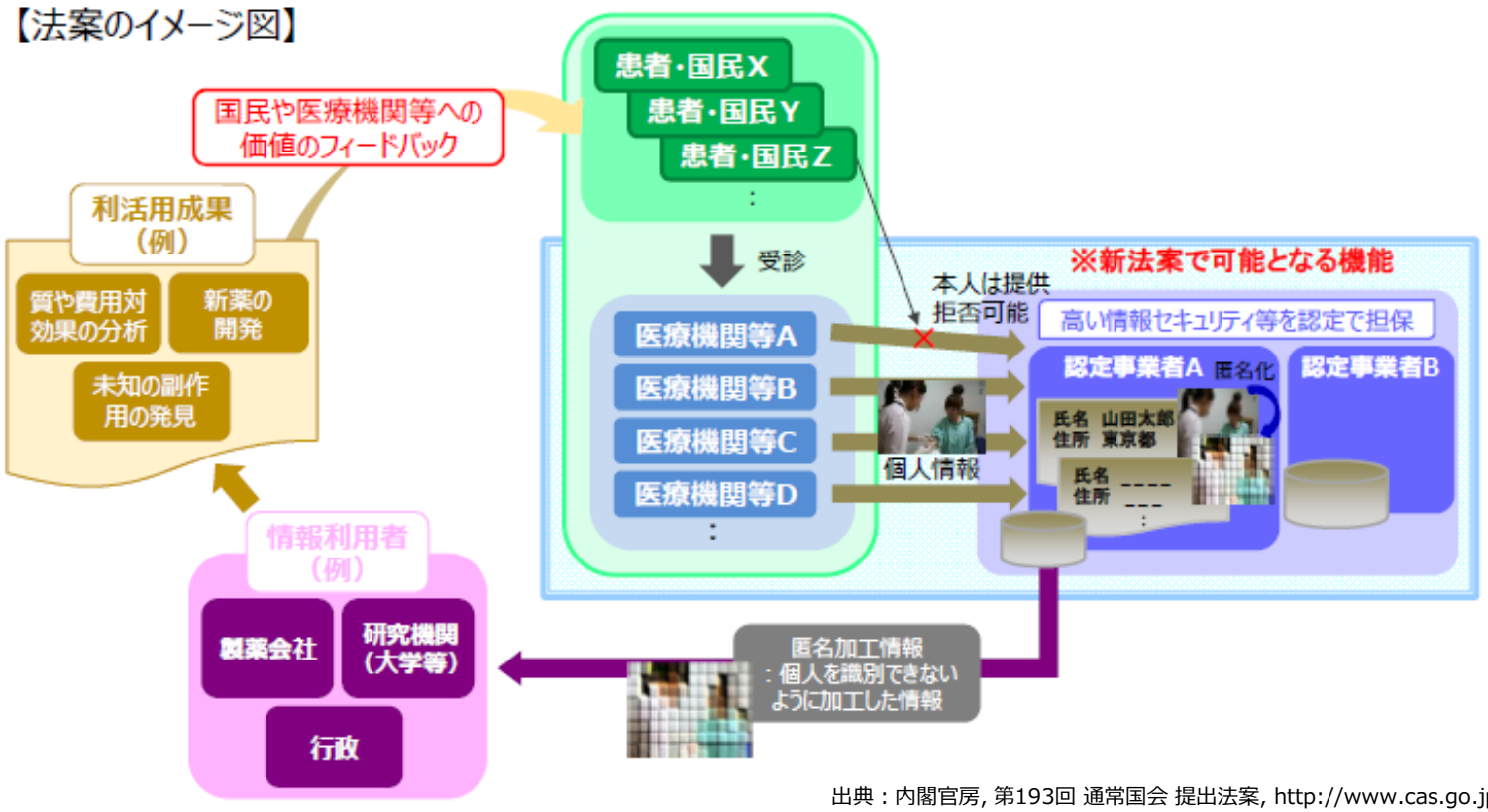
委託契約先で突合処理は禁止



※同意を取るなどが必要

医療分野における法制度：次世代医療基盤法

次世代医療基盤法※1が成立し、一定の基準を満たした「認定匿名化加工医療情報作成事業者」に対して、医療情報を収集・結合し、匿名化※2して提供可能となる※3



出典：内閣官房, 第193回 通常国会 提出法案, http://www.cas.go.jp/jp/houan/193.html

※1 「医療分野の研究開発に資するための匿名加工医療情報に関する法律案」（次世代医療基盤法案）2017年4月可決、5月公布

※2 次世代医療基盤法は「匿名加工情報」ではなく「匿名加工医療情報」である。詳細は未定。

※3 あらかじめ本人に通知することや、提供先では他の情報との照合は禁止されることなどの条件がある

安全なデータ結合分析による社会価値創造

医療情報に限らず、組織が保有する様々な機密データを、**相互に開示せずに結合した分析**を実現し、組織を越えたデータ活用による新たな知見を獲得

疾病と運動の
相関分析による
予防医療

政府・自治体

ゲノムと投薬の
相関分析による
個別化医療

医療研究機関

金融情報の
結合分析による
不正送金検知

金融機関

・・・・・

複数組織のデータを
結合した分析結果だけを提供

秘密計算技術
(安全なデータ結合)

複数組織の機密データを相互に開示せずに結合

投薬・疾病履歴、
ゲノム情報など

医療機関

運動量情報

ヘルスヘア企業

銀行残高
送金情報など

金融機関

位置情報
購買情報など

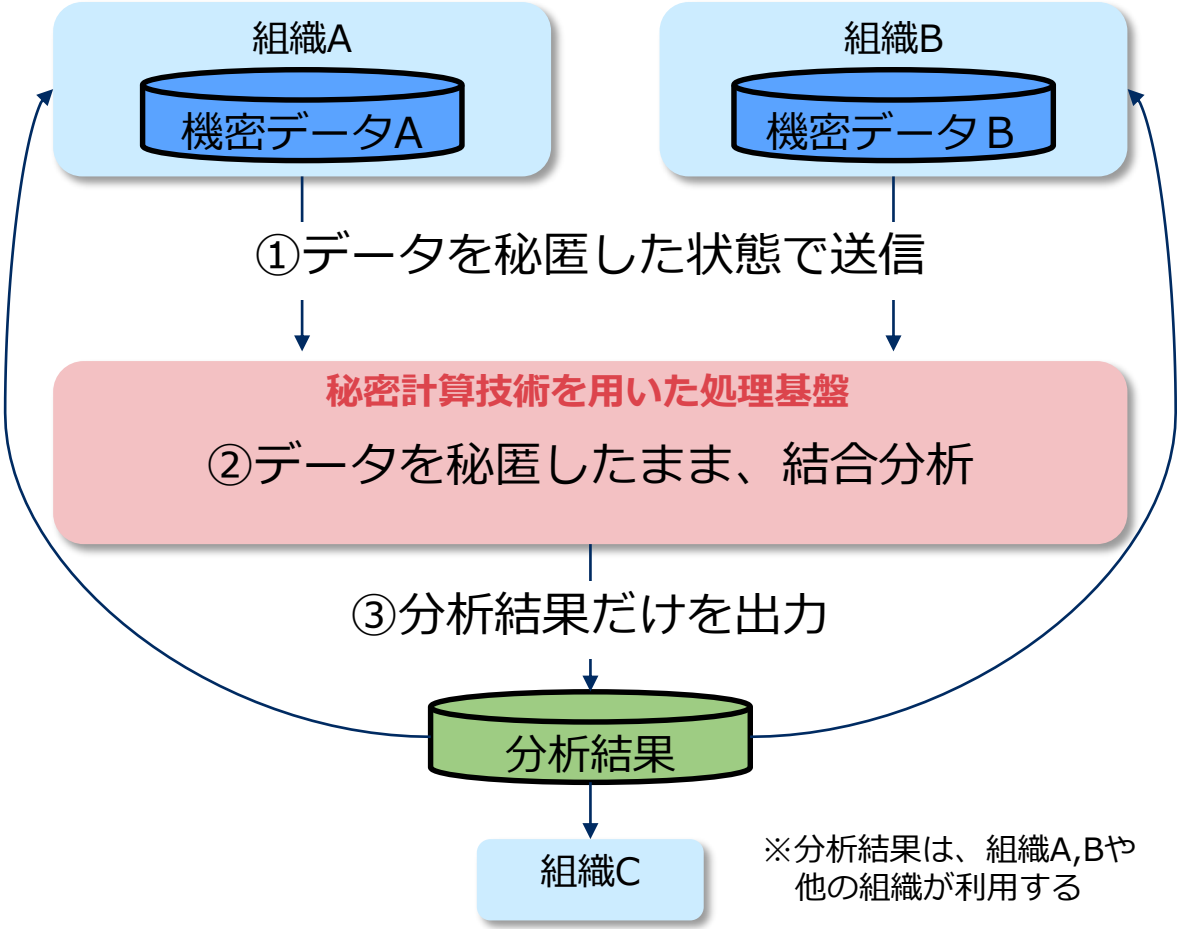
キャリア、ECサイトなど

・・・・・

2.秘密計算技術とは

秘密計算技術の概要

- 秘密計算技術とは、データを秘匿したまま処理できる技術
- 異なる組織のデータを、組織外に元データを一切開示せずに、データを結合した分析が可能

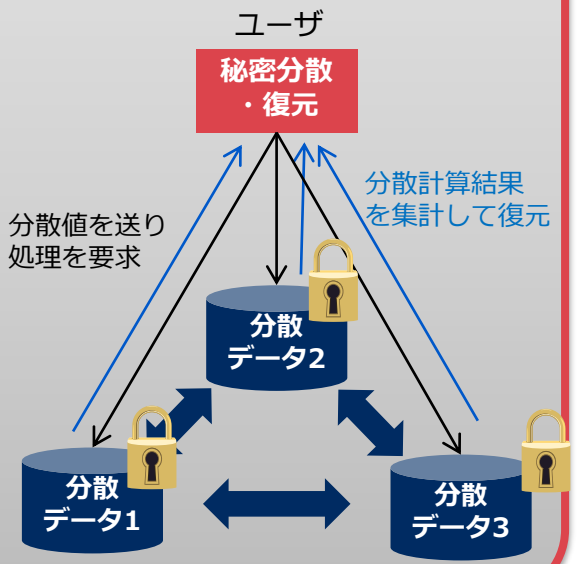


秘密計算技術の様々な方式　（本発表における用語整理）

秘密計算

秘密分散を利用した方式

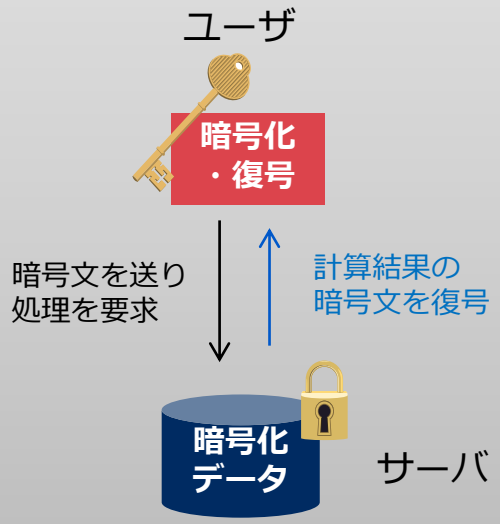
データを秘密分散したまま処理



(国内では)
比較的知られていないため、本日内容をご説明

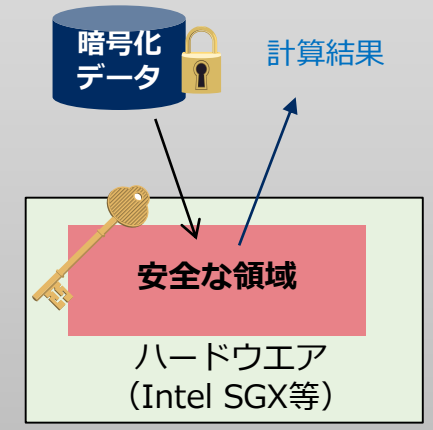
準同型暗号を利用した方式

データを暗号化したまま処理



ハードウエアを利用した方式 (Trusted Execution Environment等)

ハードウエア上の安全な領域で処理



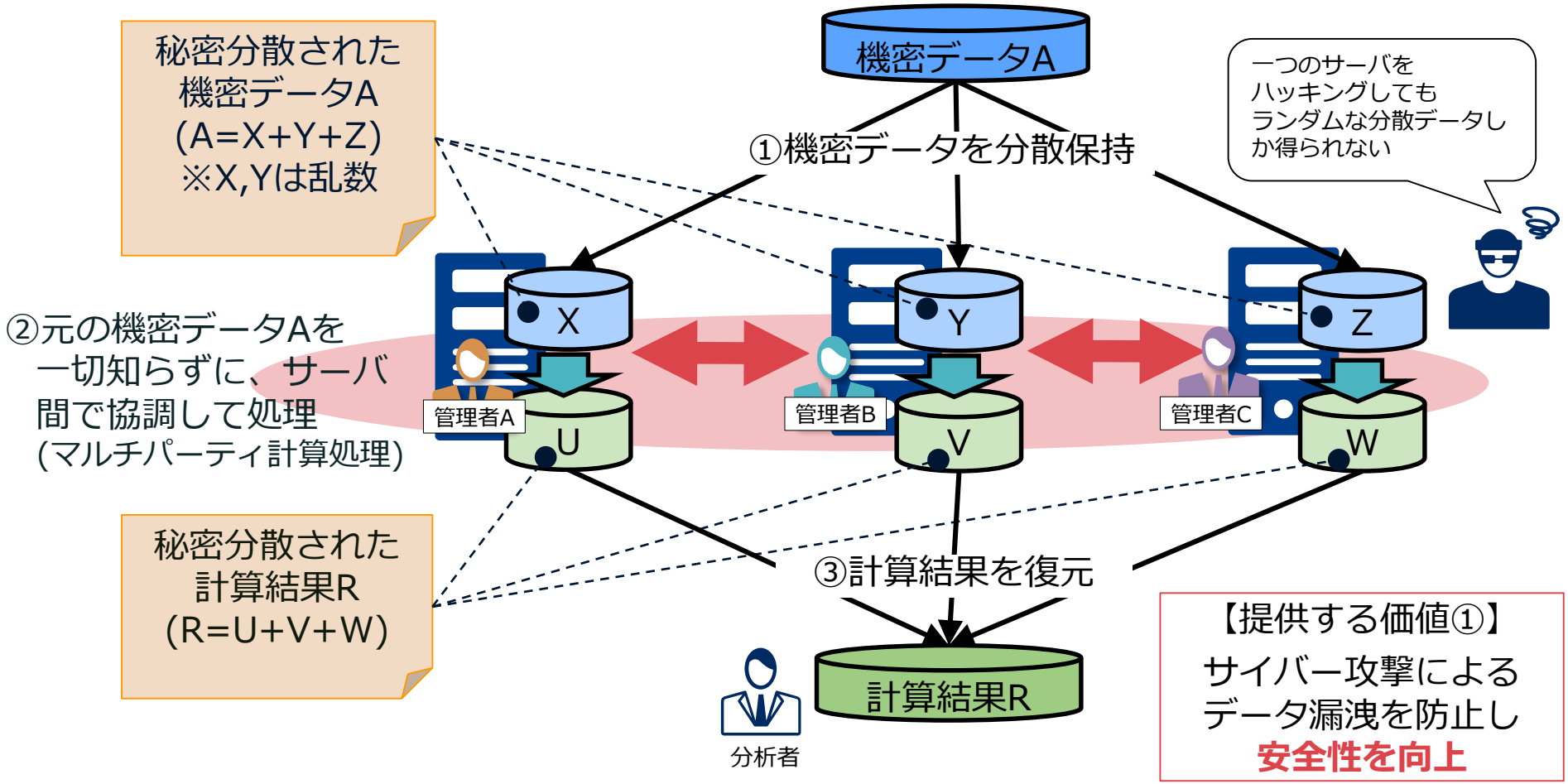
安全な領域に保持している鍵で、データを復号して処理する

・・・ その他方式も存在

秘密分散方式の秘密計算技術の説明（1/2）

秘密計算技術（秘密分散型のマルチパーティ計算技術）によって、複数のサーバが秘匿したデータを分散保持し、**秘匿したままの任意の計算**を実行可能※1

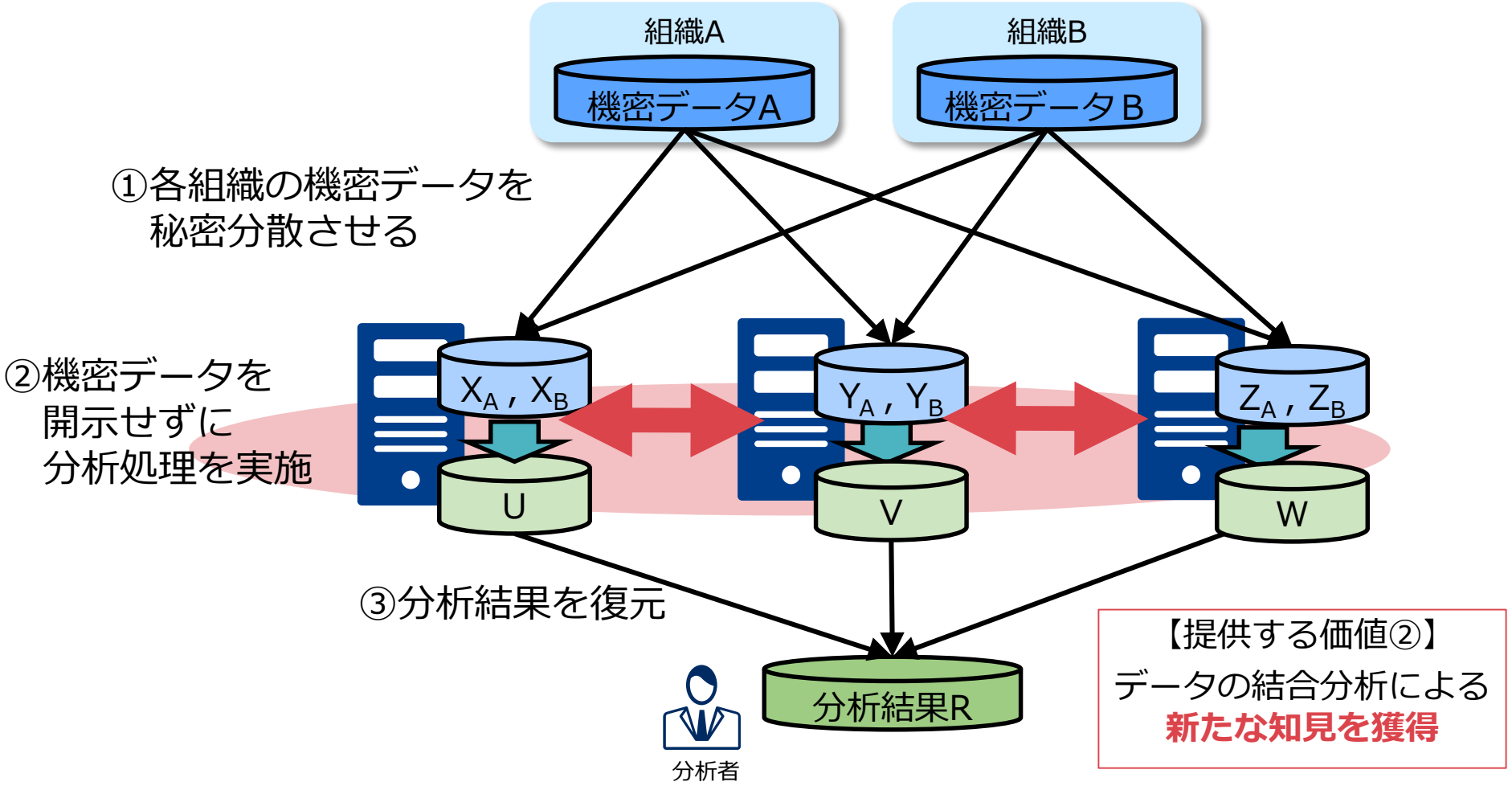
※1 計算を論理回路として表現することで、理論上は任意の計算が可能



【提供する価値①】
サイバー攻撃による
データ漏洩を防止し
安全性を向上

秘密分散方式の秘密計算技術の説明（2/2）

異なる組織のデータを秘密分散したまま処理することで、組織外に
元データを一切開示せずに、データを結合した分析結果を得る事が可能

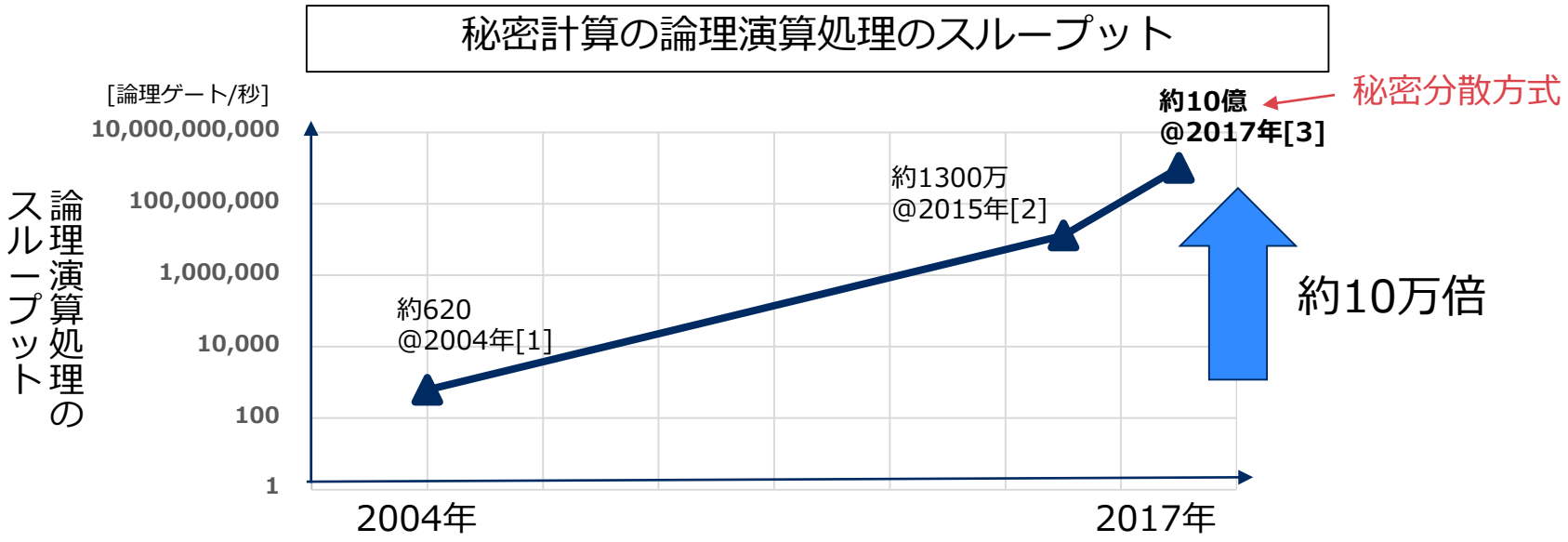


【提供する価値②】
データの結合分析による
新たな知見を獲得

秘密計算技術の性能向上

秘密計算は近年桁違いに性能が向上 ⇒ 実用可能性が高まった

例：論理演算処理は2004年以降で約10万倍の高速化※1



例：1000万件規模のデータ結合処理を数分程度で実行可能※2

- 突合処理: 約330秒 (垂直結合, 1000万件, 5属性)
- 統計処理: 約80秒 (数量表+出力の保護処理, 1000万件)
- 匿名化処理: 約15秒（1000万件, 1属性）

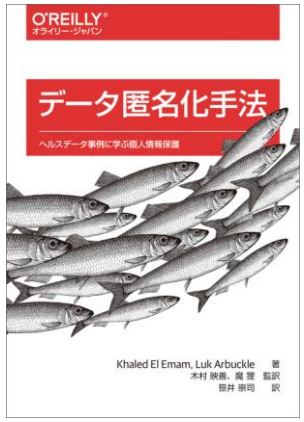
※1 詳細は「荒木 他,"秘密計算の実用可能性", SCIS2018」を参照
※2 数値はNTTの秘密計算システムを利用した場合の参考値

[1] Dahlia Malkhi, Noam Nisan, Benny Pinkas, and Yaron Sella, "Fairplay - secure two-party computation system", USENIX 2004.
[2] Shay Gueron, Yehuda Lindell, Ariel Nof, and Benny Pinkas. "Fast garbling of circuits under standard assumptions", ACM CCS 2015.
[3] Toshinori Araki, Assi Barak, Jun Furukawa, Tamar Lichter, Yehuda Lindell, Ariel Nof, Kazuma Ohara, and Adi Watzman. "Optimized honest-majority mpc for malicious adversaries - breaking the 1 billion-gate per second barrier". S&P 2017.

3.秘密計算技術を用いた安全なデータ結合

参考資料：

“データ匿名化手法 ヘルスデータ事例に学ぶ個人情報保護”
著：Khaled El Emam, Luk Arbuckle,
監訳：木村映善, 魔狸
訳：笹井崇司.
12章「セキュアな連結」

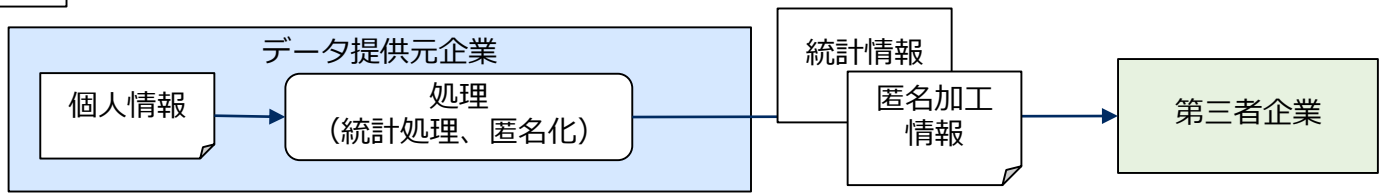


出典： オライリー社HP https://www.oreilly.co.jp/books/9784873117249/

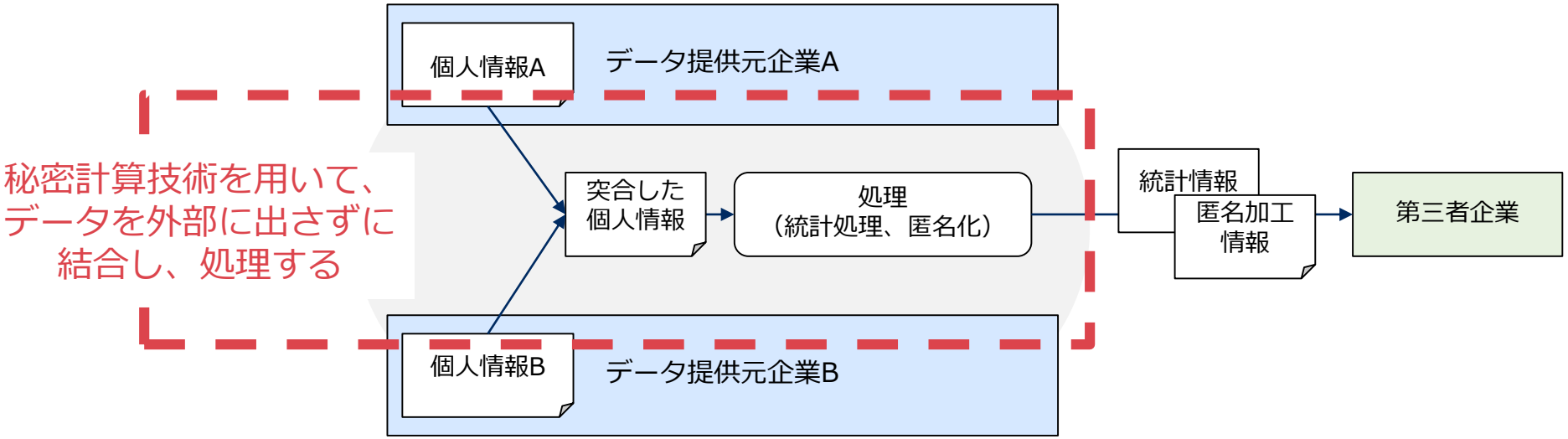
秘密計算技術による安全なデータ結合

データ提供元の複数企業が、秘密計算技術を用いることで、**責任をもって安全に個人情報を結合**して処理(統計処理、匿名加工等)を行う

現状：単一組織



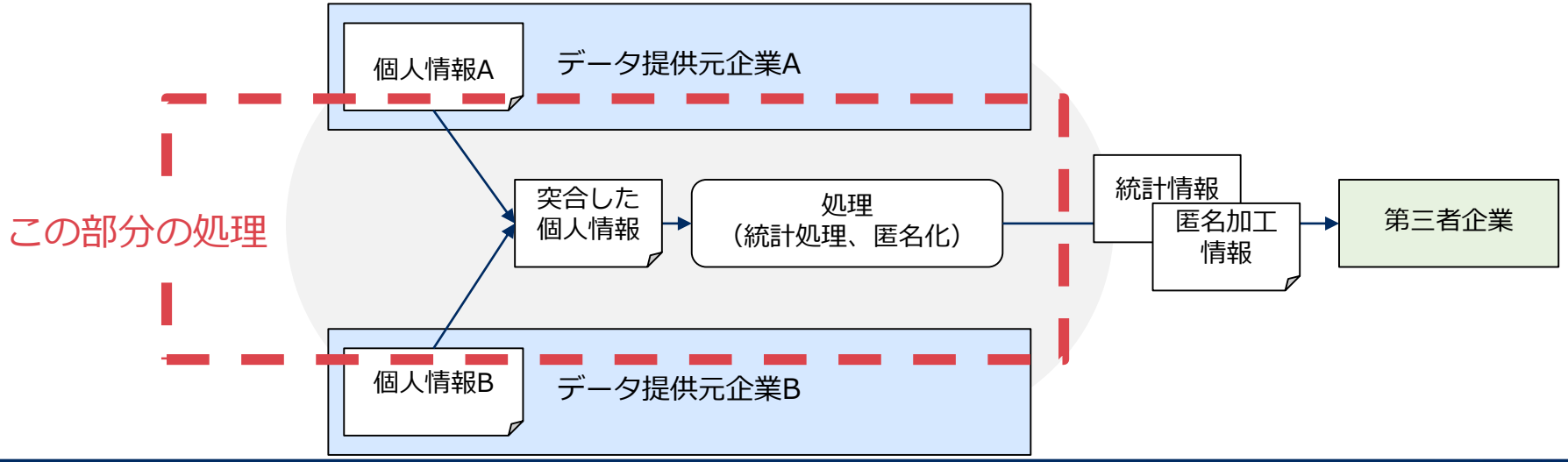
提案：複数組織



セキュアなデータ結合を実現する方式・技術を整理

実現方式の整理

秘密計算技術の利用	秘密計算技術の方式	実現方式
利用しない (完全に信頼できる機関を設置し、生データで処理)	ー	方式１
利用する	準同型暗号を利用	方式２
	秘密分散ベースを利用	方式３
	その他（本発表では割愛）	・・・

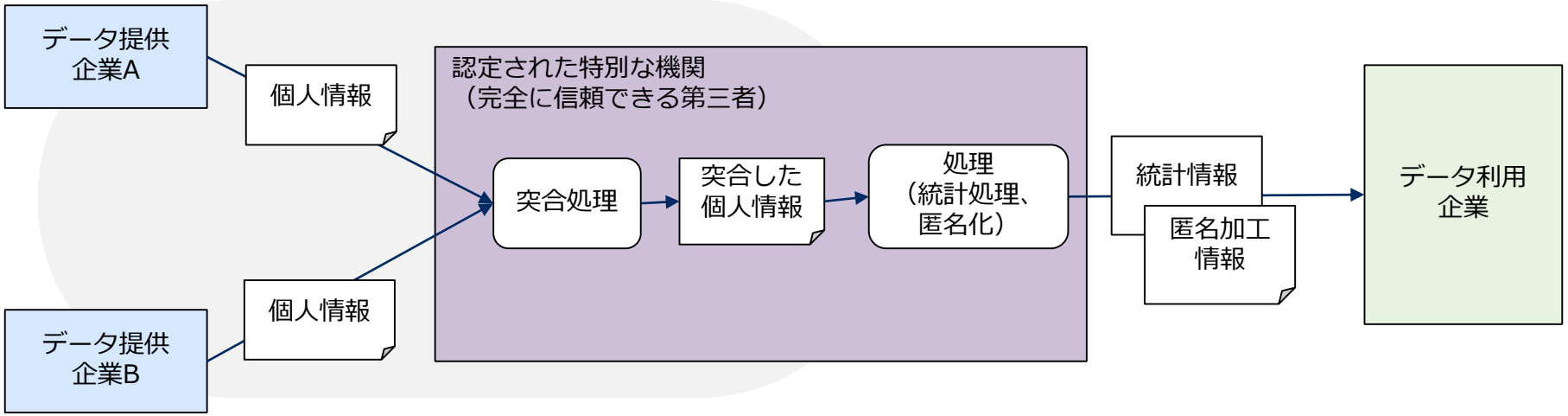


方式１：完全に信頼できる第三者を設置

認定された特別な機関にデータを送り、処理した結果を提供

- 次世代医療基盤法の”認定事業者”に近い考え

※注意： 次世代医療基盤法は「匿名加工情報」ではなく「匿名加工医療情報」である。詳細は未定。



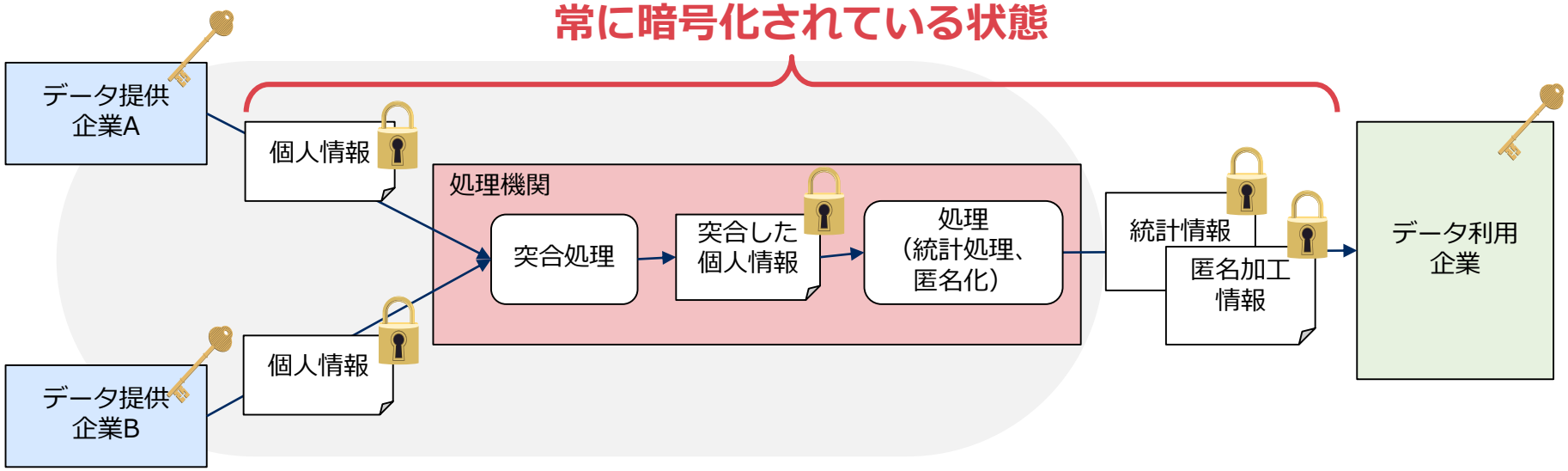
懸念点

- 「完全に信頼できる第三者」は生データを閲覧可能
- 「完全に信頼できる第三者」の認定・監査等のコストが大きい（「データ匿名化手法」本の主張）
- 一部の機関だけが認められると、処理内容（統計処理等）の多様化が望めない

方式2：準同型暗号を利用した例

準同型暗号：暗号化したまま、復号せずに処理できる技術
- 例： Enc(a) × Enc(b) = Enc(a+b)

データ提供元で暗号化し、暗号化したまま処理し、処理結果を復号する



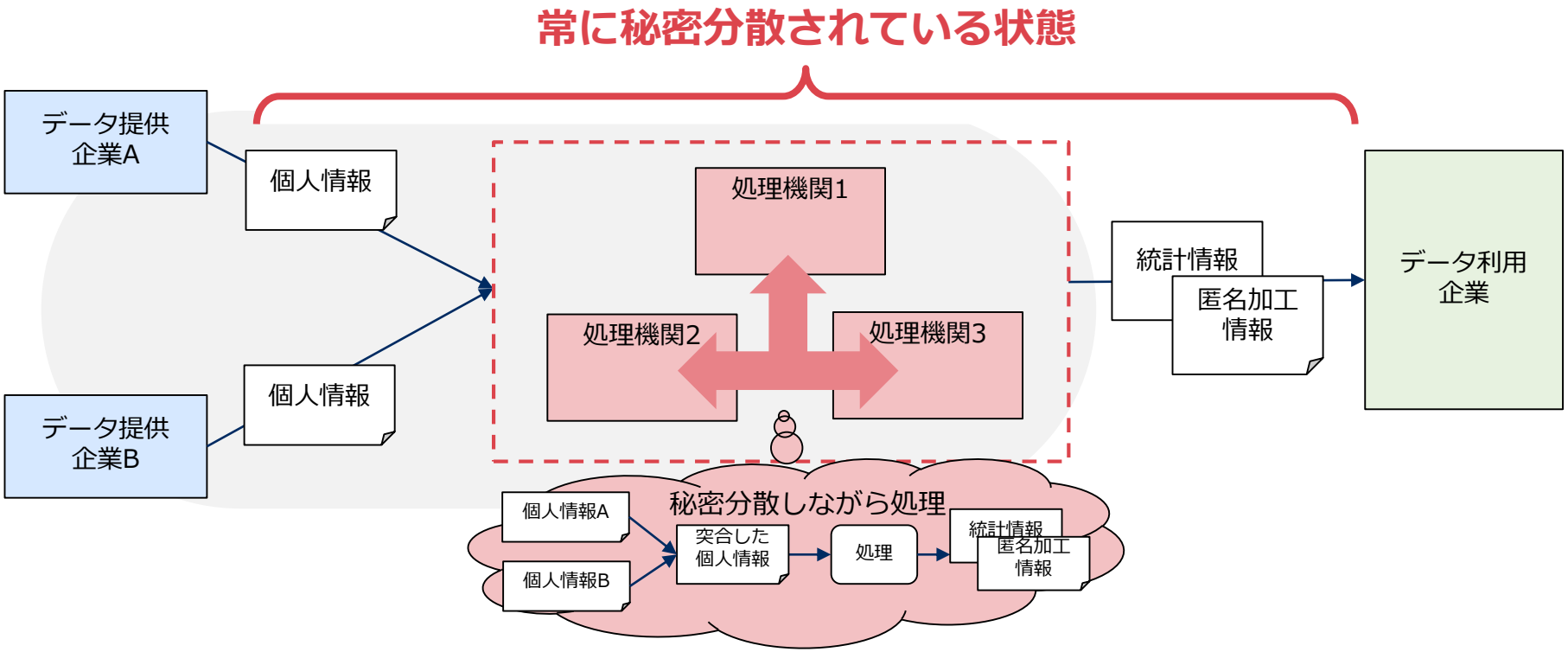
メリット：生データが閲覧可能な「完全に信頼できる第三者」の設置が不要

安全であるための条件：
- 鍵の適切な管理（例：不正入手した鍵によって処理途中での復号できてしまう）
- 結託防止（例：暗号化したデータを鍵を持つ他者へ提供できてしまう）

⇒ 鍵管理の第三者の設置、鍵の安全管理義務、結託禁止規定などが必要？

方式３：秘密分散ベースの秘密計算を利用した例

- 秘密分散ベースの秘密計算を用いて、複数サーバに秘密分散させながら、個人情報を結合した処理を実施
  - 例：3者の処理機関で秘密計算を実施する例



- メリット：信頼できる第三者の設置が不要
- 安全であるための条件：結託防止（例：結託すれば処理途中のデータを取得可能）
  ⇒ 結託禁止の規定などで制限できないか？

まとめ

本発表で紹介した3方式の整理

	説明	運用コスト	技術のコスト
方式１	完全に信頼できる機関で実現	大	低 （生データ処理するため）
方式２	準同型暗号方式の秘密計算で実現	小	中～大 （処理や安全性に依存するが、生データ処理より高コスト）
方式３	秘密分散方式の秘密計算で実現	小	

- 完全に信頼できる第三者の設置は、認定・内部外部監査等の運用コストが懸念 （「データ匿名化手法」本の主張）
- 技術コストは、今後の技術発展で許容できる範囲になる可能性あり

安全性の前提

- 暗号化： 鍵の安全管理など
- 秘密分散： 結託の防止など

⇒ 前提を担保する制度にするのは？ 例：匿名加工情報と他データとの突合禁止の条項

4. 制度の方向性・まとめ

制度整備に関しての検討

企画セッションで出た案 （板倉先生の資料より抜粋）

- 法律（個人情報保護法）の改訂
  - » 何条のどこの文言を改正するのか？
- 個人情報保護委員会の規則（個人情報保護法施行規則）の改訂
  - » 19条1号から5号の解釈なのか？
  - » 1号から5号以外に条項が必要なのか？
- 個人情報保護委員会のガイドライン（匿名加工情報編）の改訂
  - » 1号から5号の解釈として示すのか？
- 認定個人情報保護団体の個人情報保護指針（例：JIPDEC）
  - » 匿名加工情報について独自の解釈を（届出を通じて）導入させることができるのか？
- 自治体の条例
- 法解釈
  - » 例：情報法制研究所のパブコメ

改訂までに
時間を要する

活動の案

- **法解釈・改正の可能性検討が必要**
  - ⇒ **法学者・政府関係者・研究者を交えて議論していきたい**
- **実績・実例が必要**
  - ⇒ **サンドボックス/特例などの活用の可能性を検討**

まとめ

- 秘密計算技術によって、複数組織がもつデータを
  安全に結合して処理が可能
- 秘密計算技術の方式
  - 準同型暗号を用いた方式
  - 秘密分散を用いた方式
  - など
- 社会実装には、制度整備や実証などが必要

技術や制度を議論する勉強会を行う予定です。
是非皆さんと議論をさせてください。