資料4－1

Innovative R&D by NTT

近未来におけるパーソナルデータ活用のための技術

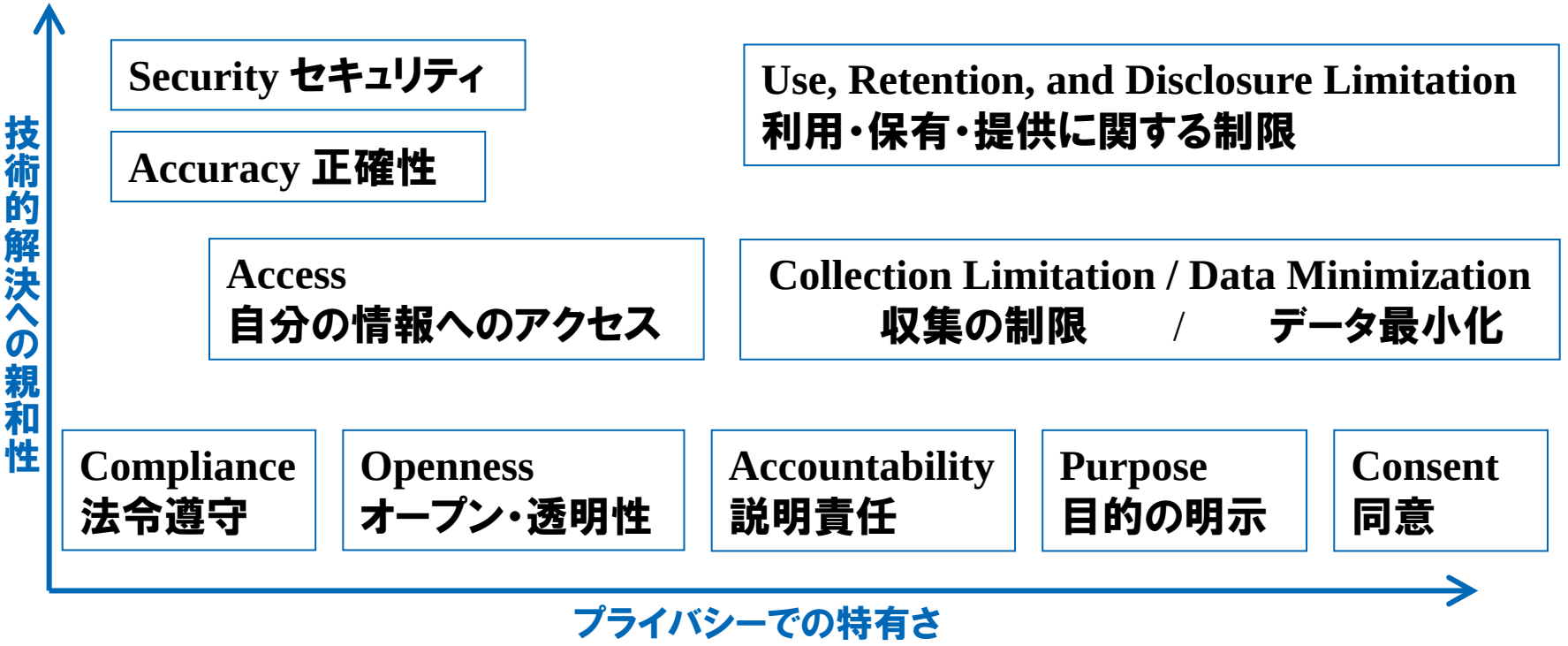
2015.7.9

日本電信電話株式会社
NTTセキュアプラットフォーム研究所
高橋克巳

パーソナルデータのプライバシー原則

パーソナルデータのプライバシー原則

- パーソナルデータ管理者が守るべき原則が掲げられて来た
- 日本の個人情報保護法も本原則を踏襲している



＊ Privacy Principles from “Creation of a Global Privacy Standard” Ann Cavoukian (2006)
ISO/IEC 29100 Privacy frameworkも参考にした（10原則の分類・配置は筆者による）

これまでのパーソナルデータ

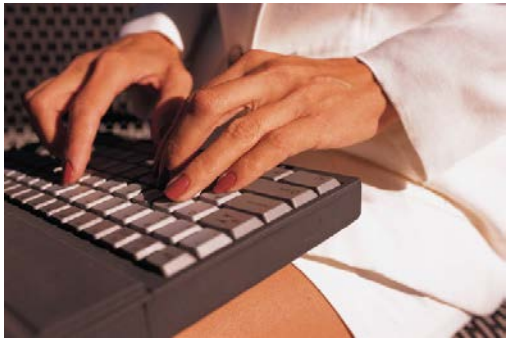
これまでのパーソナルデータの例

- 「書面」で取得されるもの
  - 氏名、住所、生年月日・・・
- サービス提供に伴って取得されるもの
  - 買い物履歴
  - 移動履歴
    - 鉄道乗降、携帯位置情報
  - Web閲覧履歴

→ 理解された手段で、正しい目的で

これまでのパーソナルデータの使い方の例

パーソナル
データ



分析者

業務改善

一つの会社内

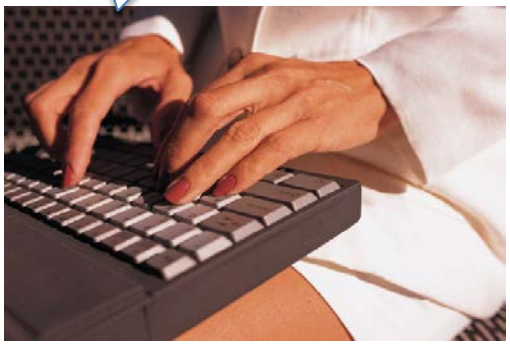
この使い方は？

匿名加工情報（H27 改正個人情報保護法）

匿名加工して

パーソナル
データ

規律の元で



分析者

業務改善

鉄道

- 乗降履歴 → 仕入れ数量決定（できる）

飲食店

- 乗降履歴 → 自宅住所判定　（できない）

再特定禁止の規律

匿名情報の分類

十分な匿名情報

年齢	購入品
40代	食品，スポーツ用品
40代	食品，スポーツ用品

	食品	スポーツ用品
30代	10人	5人
40代	23人	12人

（統計情報）

匿名加工情報

仮名	年齢	購入品
ID001	41	パン，牛乳，野球用品，野球用品，キャップ，・・・
ID002	48	おにぎり，シャツ，スパイク，不動産，・・・

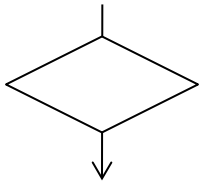
※ 加工の度合いには様々なレベルがある

個人情報

氏名	年齢	購入品
鈴木二郎	41	パン，牛乳，バット，ボール，キャップ，・・・
三浦数良	48	おにぎり，シャツ，スパイク，古城，・・・

匿名化技術とは何をする技術なのか？

会員番号、生年月日、住所、年齢、購買品1、購買品2、購買品3、．．．．．



そのまま用いていいのか、
加工すれば大丈夫か、削除するのか

会員番号、生年月日、住所、年齢、購買品1、購買品2、購買品3、．．．．．

識別子（削除）　準識別子（加工）　非識別子（そのまま）

- 匿名化技術は、取り決めに従って属性に対して、削除、加工、無加工のどれかの操作を行うこと
- 活用の際、どの属性をどのように扱うのかを取り決めるのが、個人情報取り扱い責任者の責務
- 非保護の属性の選定には十分な注意が必要であり、準識別子との境界は明確ではない

※ 本事例は例であって、加工して保護するのはは住所と年齢のみであればよいという意味ではない
※ 識別子を削除、準識別子を加工し、非識別子をそのまま用いるという従来の考え方はビッグデータでは問題があると考える必要がある

k-匿名性（同じ属性を持つ人が「k人」以上いるようにする）

Innovative R&D by NTT

会員番号	生年月日	住所	年齢	購買品
1001	1979.04.01	東京都中央区A町	34	パン、ガム、新聞、…
1002	1986.12.10	神奈川県横浜市A町	26	鉛筆、弁当、漫画、…
1003	1974.10.10	東京都渋谷区B町	38	ガム、アイス、チョコ、…
1004	1991.05.05	神奈川県鎌倉市B町	22	書籍、新聞、電池、宝石、…
1005	2006.11.10	埼玉県川越市A町	17	化粧品、あめ、アイス、…
1006	1990.02.06	神奈川県厚木市C町	23	時刻表、鉄道模型、カメラ、…
1007	2003.08.15	埼玉県浦和市B町	19	ネジ、ビス、ハンマー、…
1008	2000.09.30	埼玉県大宮市C町	9	肉まん、ガム、新聞、…
1009	1983.01.01	東京都練馬区C町	30	コーラ、弁当、雑誌、…
1010	1994.07.07	埼玉県与野市D町	18	ガム、水、ドリンク剤、…

削除　　加工（保護）　　そのまま（非保護）

k-匿名性（k=3）を満たした状態

会員番号	生年月日	住所	年齢	購買品
1001	1979.04.01	東京都	30代	パン、ガム、新聞、…
1003	1974.10.10	東京都	30代	ガム、アイス、チョコ、…
1009	1983.01.01	東京都	30代	コーラ、弁当、雑誌、…
1002	1986.12.10	神奈川県	20代	鉛筆、弁当、漫画、…
1004	1991.05.05	神奈川県	20代	書籍、新聞、電池、宝石、…
1006	1990.02.06	神奈川県	20代	時刻表、鉄道模型、カメラ、…
1005	2006.11.10	埼玉県	未成年	化粧品、あめ、アイス、…
1007	2003.08.15	埼玉県	未成年	ネジ、ビス、ハンマー、…
1008	2000.09.30	埼玉県	未成年	肉まん、ガム、新聞、…
1010	1994.07.07	埼玉県	未成年	ガム、水、ドリンク剤、…

3

3

4

NTT

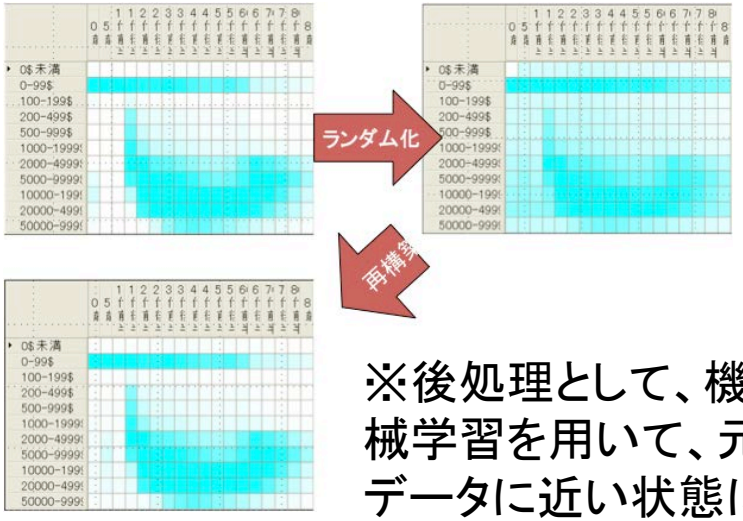
Pk-匿名性

その人が誰であるか 1/k 以上の確率で当てられないようにする

会員番号	生年月日	住所	年齢
1001	1979.04.01	東京都中央区A町	34
1002	1986.12.10	神奈川県横浜市A町	26
1003	1974.10.10	東京都渋谷区B町	38
1004	1991.05.05	神奈川県鎌倉市B町	22
1005	2006.11.10	埼玉県川越市A町	6
1006	1990.02.06	神奈川県厚木市C町	23
1007	2003.08.15	埼玉県さいたま市B町	9
1008	2000.09.30	埼玉県川口市C町	12
1009	1983.01.01	東京都練馬区C町	30
1010	1994.07.07	埼玉県与野市D町	18

削除　　保護（ランダム化）

会員番号	生年月日	住所	年齢
1001	1979.04.01	神奈川県厚木市C町	30
1002	1986.12.10	神奈川県横浜市A町	26
1003	1974.10.10	東京都渋谷区B町	38
1004	1991.05.05	神奈川県鎌倉市B町	24
1005	2006.11.10	埼玉県川越市A町	6
1006	1990.02.06	東京都中央区A町	23
1007	2003.08.15	埼玉県さいたま市B町	9
1008	2000.09.30	埼玉県川口市C町	12
1009	1983.01.01	東京都練馬区C町	34
1010	1994.07.07	埼玉県与野市D町	18



※後処理として、機械学習を用いて、元データに近い状態に戻す推定を行う（再構築）

近未来のパーソナルデータ

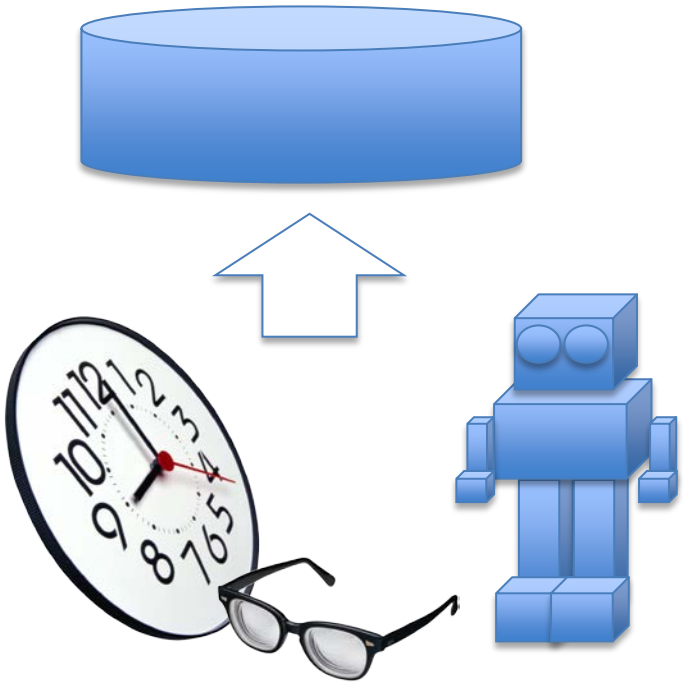
近未来のパーソナルデータの例（その１）

- 機器が記録し続ける
  - 時計が：GPS、加速度センサー、心拍センサー、…
  - 家電が：テレビ視聴、冷蔵庫開閉、トイレ利用、…
- 会話は残り続ける
  - SNS や chat でのおしゃべりが
  - 対話型ロボットが自宅にいつでもいる

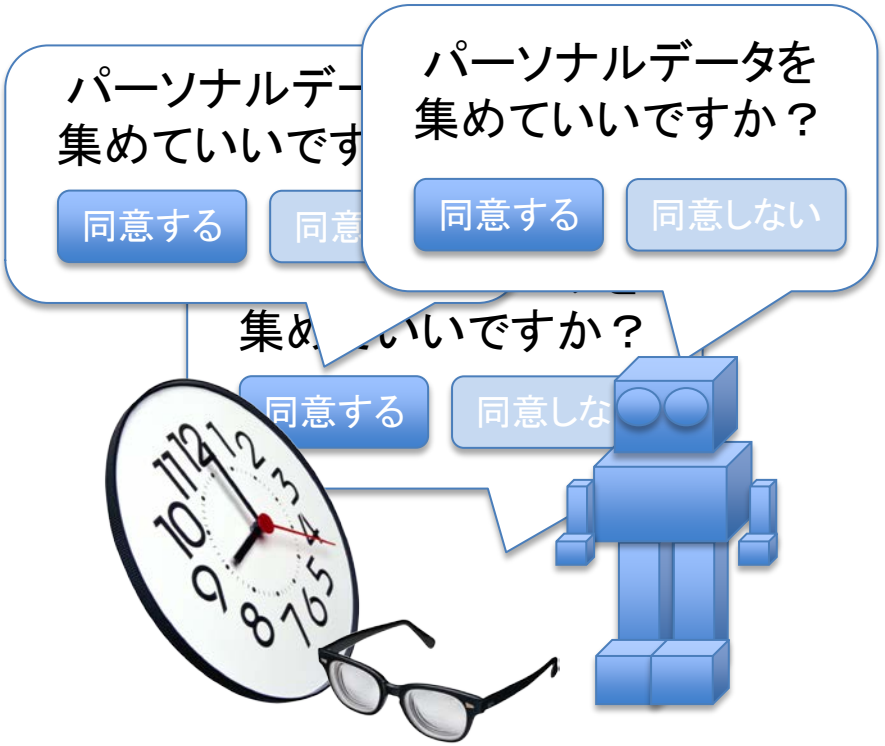
→ 知らぬ間に提供？／いちいち確認？

パーソナルデータの提供問題（どちらもやっかい）

知らぬ間に提供



いちいち確認



- 提供可否をその都度判断する《エージェント》が必要
  - データ提供の《影響》判断は極めて高度なスキル
  - 最終的には提供先との《トラスト》の問題

近未来のパーソナルデータの例（その２）

Innovative R&D by NTT

A.《横に長い》ビッグデータ化する

1. 時系列が長い
   - ■■■■■■■■■■■■■■■……
   - １日分の健康データ　→　５年分の健康データ

2. オープンデータと統合されて長くなる
   - ■■■■■■■＋○○○○○○
   - ある人のデータ　＋最寄小学校名，近隣コンビニ数，…

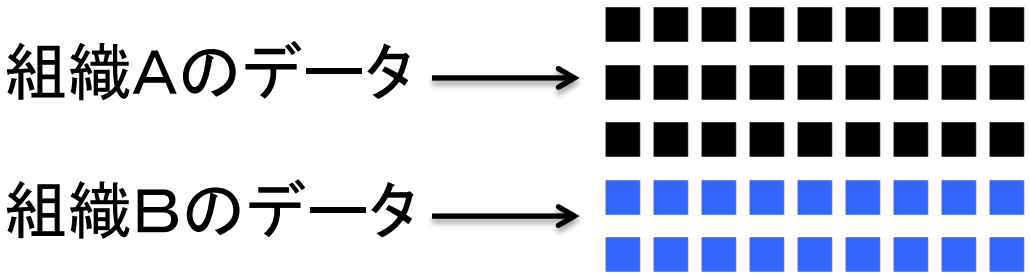
3. 他のパーソナルデータと統合されて長くなる
   - ■■■■■■■＋△△△△＋◇◇◇
   - ある人のデータ　＋移動履歴＋購買履歴，…

NTT

近未来のパーソナルデータの例（その２ 続）

B.《人数が多い》ビッグデータ化する

1. 他のパーソナルデータと統合されて長い

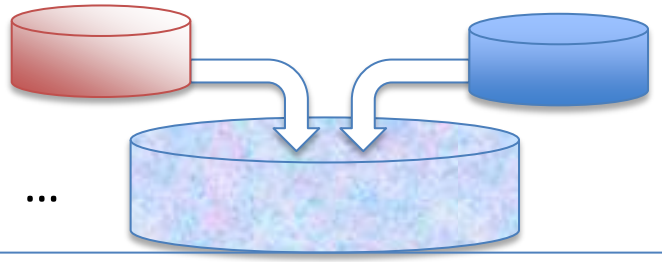


パーソナルデータのビッグデータ化問題

Innovative R&D by NTT

A-3　他のパーソナルデータと統合

- ■■■■■■■ ＋△△△△＋◇◇◇
- ある人のデータ ＋移動履歴＋購買履歴，…



- プライバシーの問題
  - データ最小化、利用・保有・提供に関する制限、オープン・透明性
  - 営業秘密の混合

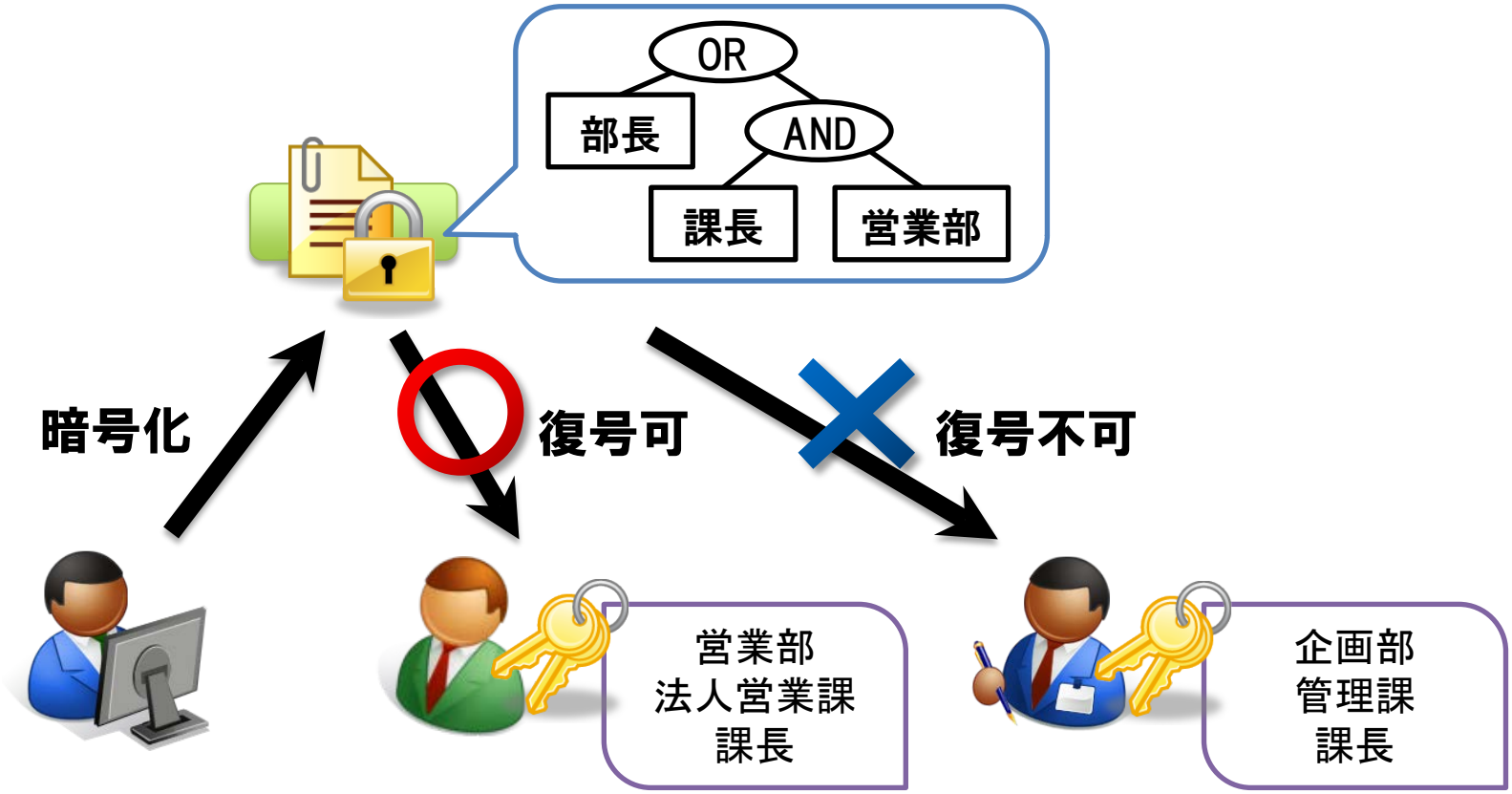
- どうすればよいのか？
  - プライバシーとビッグデータの両立
    - データは いつでも 集められる状態にしておき、必要の応じて最小限の分析を行えばよいのではないか
  - 《暗号プロトコル》（マジックプロトコル）
    - 必要な人が必要な情報だけ使えるように（→トラストの実現）

NTT

暗号プロトコル
マジックプロトコル

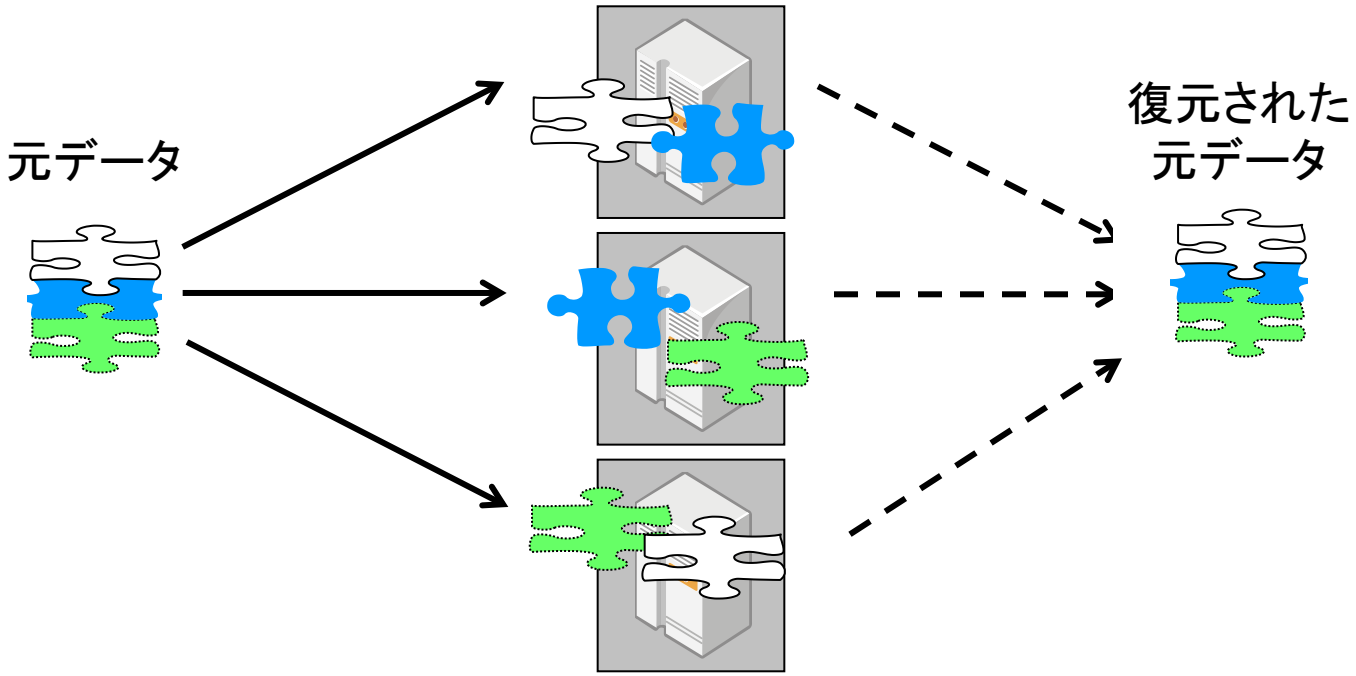
条件（人や時間）で開示を制御する暗号

- インテリジェント暗号（関数型暗号）
- 図の例
  - 秘匿したいデータに条件を設定：部長または営業課長
  - 閲覧時：条件に合致した場合のみ開示される



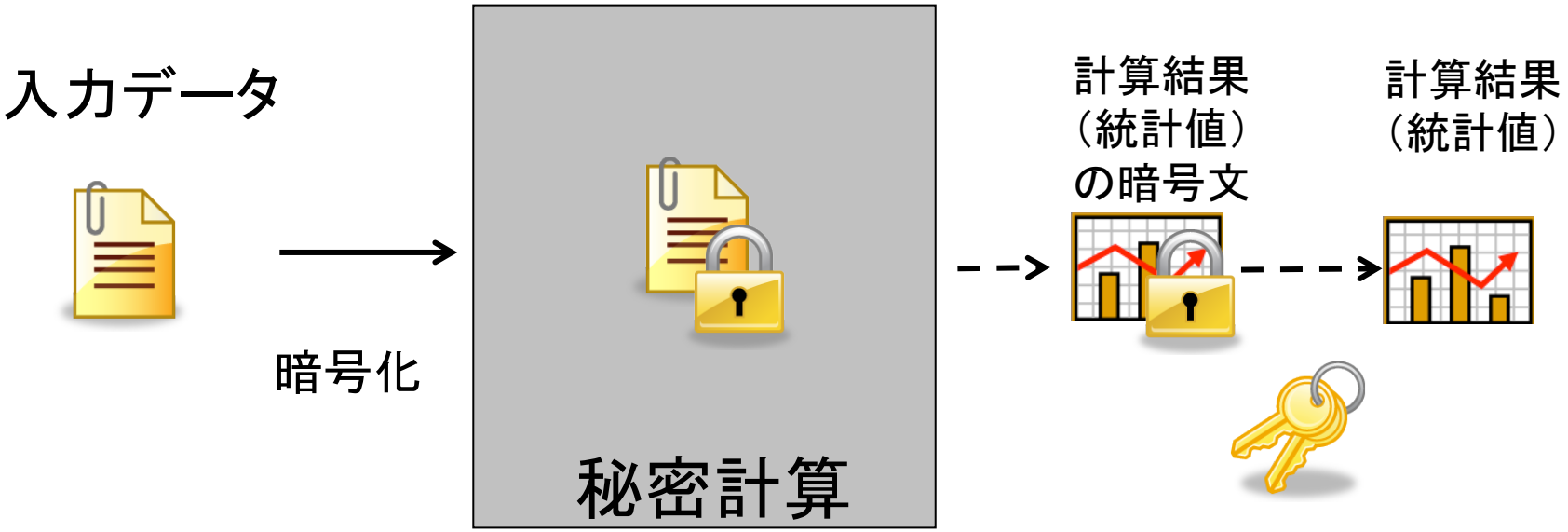
秘密分散

- 複数人の合意で開示を制御する暗号
- ２人の許可がなければ情報開示ができない
- １台盗んでも何の情報も得られない（秘匿性）
- １台故障しても残りからデータを復元できる（可用性）



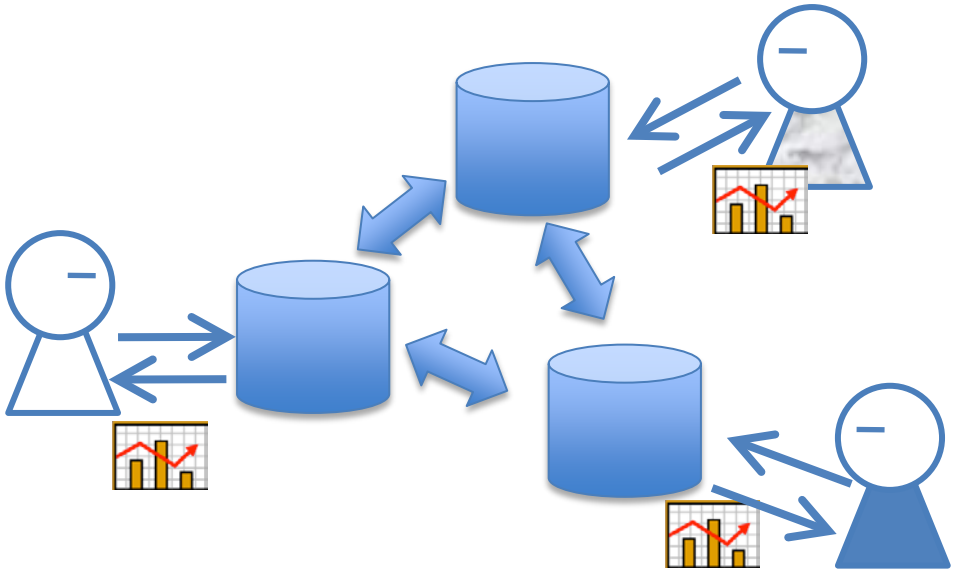
秘密計算（委託型）計算結果のみの開示する

- データを暗号化して入力、データを盗んでも何の情報も得られない
- 暗号化したままデータを計算、計算する側も何もわからない
- 計算結果の暗号文が出力（結果のみが分かる）



秘密計算（参加型）データを秘匿した共同分析

- データをそれぞれが暗号化して入力
- 暗号化されたままのデータを共同で計算（マルチパーティー計算）
- 誰も他人のデータがわからない
- 計算結果の暗号文が出力（結果のみが分かる）



ありがとうございました

