セッション2: 新型コロウイルス対策を踏まえた社会経済の変革

プライバシーに配慮したデータ利活用

国立研究開発法人情報通信研究機構
経営企画部 / サイバーセキュリティ研究所

盛合 志帆

私のPosition Statement

ウィズコロナ期こそプライバシーに配慮したデータ利活用を

- 接触確認アプリ
  - 経済活動の再開・継続を支えるツールの一つ
  - 公衆衛生上の有益性とプライバシーはトレードオフの関係
- プライバシー保護データ解析
  - プライバシーを保護したままどのようにデータ解析？

接触確認アプリのしくみ

<通常時>

- 他者との接触についてアプリの端末に**相手の識別子（個人に紐付かない）**が記録される。
- 識別子の記録は、一定期間経過後に順次削除されていく。

接触の条件をを満たしたら識別子を記録

<陽性確認時>

- 保健所で感染者システムに陽性者が登録される。
- 登録された陽性者は保健所の通知を受けて、自分が陽性者であることをアプリ上で入力。
- アプリユーザーに対して、陽性者との接触歴がある場合に**接触者アラートが通知され、これを確認。**
  （**接触した個人が特定できない形で通知**）
- 接触が確認された者には、メッセージにより、**適切な行動と帰国者・接触者相談センターへの相談方法等をガイダンス。**

参照：「接触確認アプリ及び関連システム仕様書」新型コロナウイルス感染症対策テックチーム (2020.5.26)

日本における接触確認アプリ

これまでの主な経緯

4月6日	内閣官房「新型コロナウイルス感染症対策**テックチーム**」キックオフ
4月11日	AppleとGoogle：濃厚接触の可能性を検出するアプリの**共同開発**を発表
5月8日	厚生労働省が**一元的にアプリ開発**することに決定 3月下旬よりCode for Japanなどが政府と協力して接触確認アプリの開発を進めていたが、AppleとGoogleが公衆衛生当局による開発と「1国1アプリ」をAPI利用の条件と定めたため
5月9日	「接触確認アプリに関する**有識者検討会**」発足、アプリの仕様検討
5月20日	AppleとGoogle：濃厚接触の可能性を検出するシステム「Exposure Notification System」の**APIを一般公開**
5月26日	テックチーム：接触確認アプリに関する**仕様書等の公表**

接触確認アプリの分類(各国比較)

陽性者データ管理	国が管理（中央サーバ型）		端末で管理（分散型）
個人情報の取得	個人特定型 (電話番号等)	匿名型	
代表例	カタール シンガポール オーストラリア	イギリス フランス	ドイツ スイス エストニア 日本

接触確認アプリ仕様に対する
プライバシー・セキュリティ上の評価等

- **プライバシー**
  - 感染者システムから発行される処理番号は「要配慮個人情報」であり、安全管理措置の義務を負う
  - 利用にあたってはユーザの同意を取得することを原則とする
  - 情報のライフサイクル(取得、保管、利用、移転、削除)の各過程においてプライバシーに対する十分な配慮
- **セキュリティ**
  - 「政府機関等の情報セキュリティ対策のための統一基準」に基づくセキュリティ対策
  - スマホ端末のアプリに関するセキュリティ対策
- **運用上の留意点**
  - 透明性の確保：　仕様書等の公開、ユーザーへの通知公表
  - インクルーシブネス/包摂性：多くのユーザーが利用できるデザイン
  - 使用目的の限定：目的外利用の禁止、感染症終息後のサービス停止
  - 検証：アプリ運営者の継続的な検証と有識者検討会への報告・評価

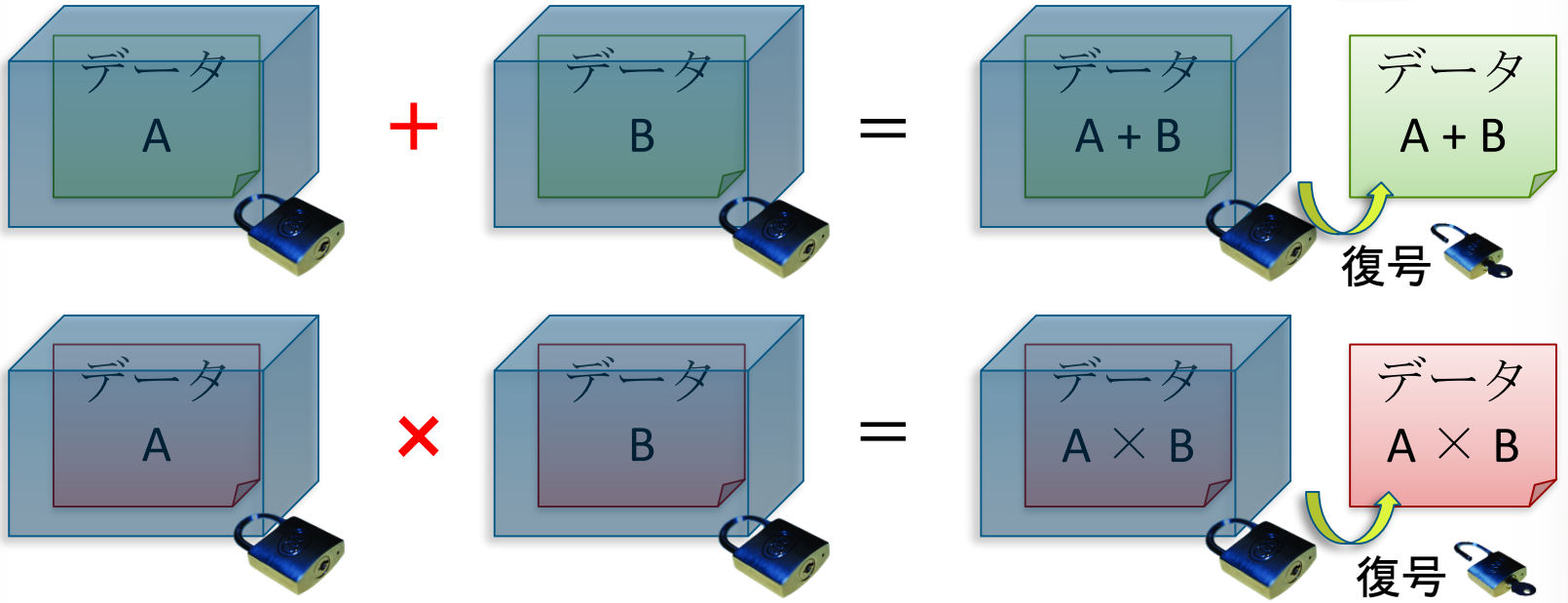
データを秘匿したまま解析を行う

プライバシー保護データ解析

プライバシー保護データ解析

じゅんどうけいあんごう
準同型暗号：
暗号化したまま計算ができる

もとのデータ
(AやB)を知らずに
和や積を計算できる

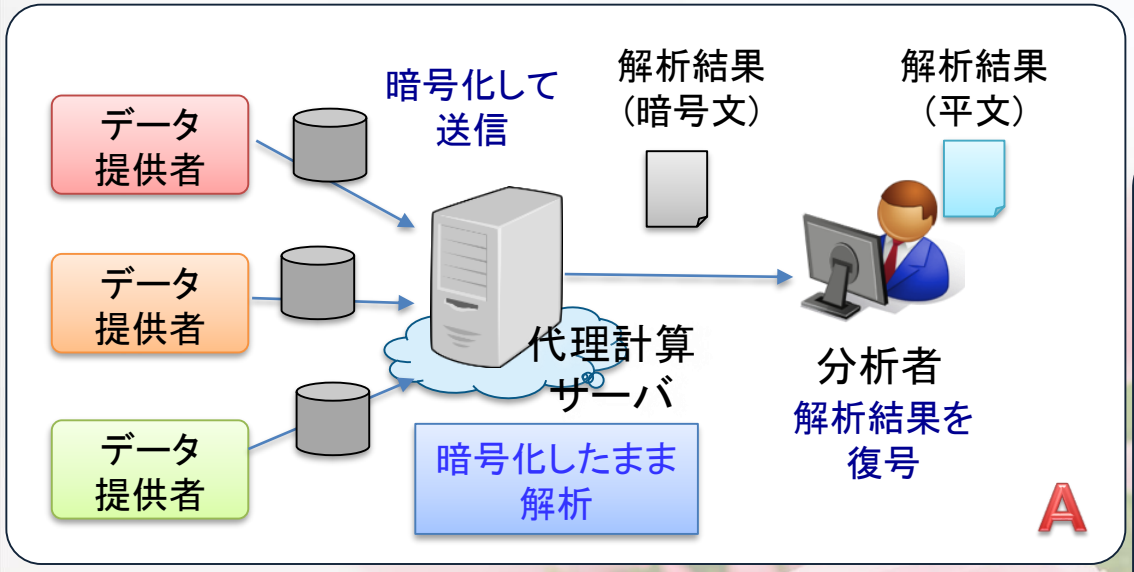


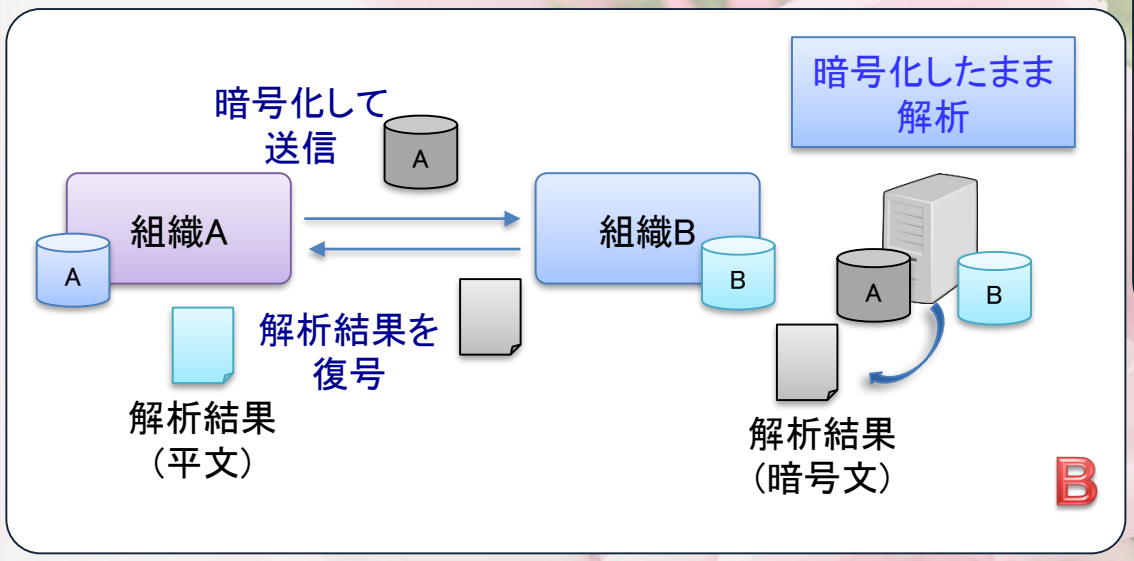
完全準同型暗号：加算、乗算の両方ができるもの

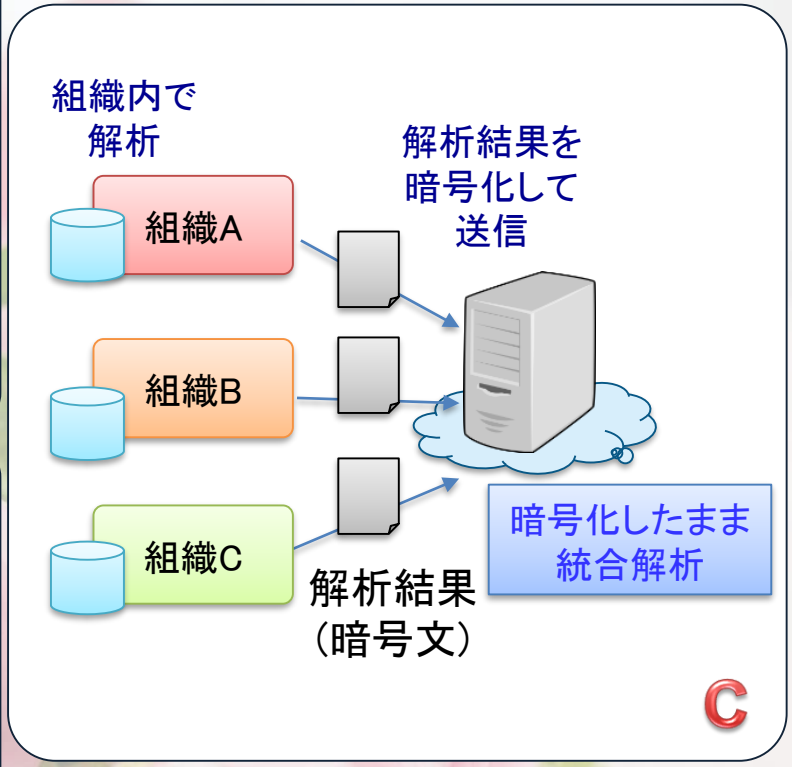
- 2009年にGentryが完全準同型暗号（格子ベース）を初めて構成
- **暗号化したままでのデータ解析**に道が開けた

プライバシー保護データ解析

利用シナリオの例

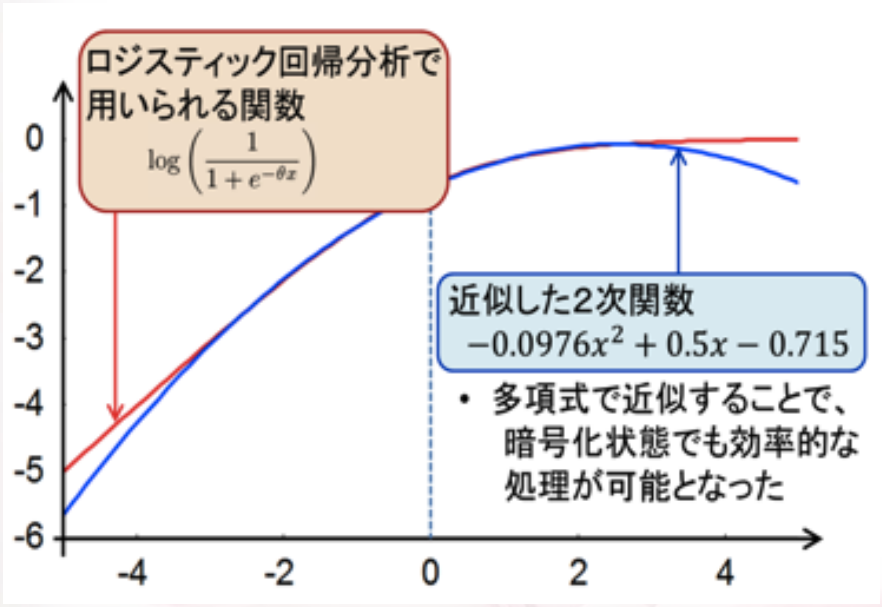


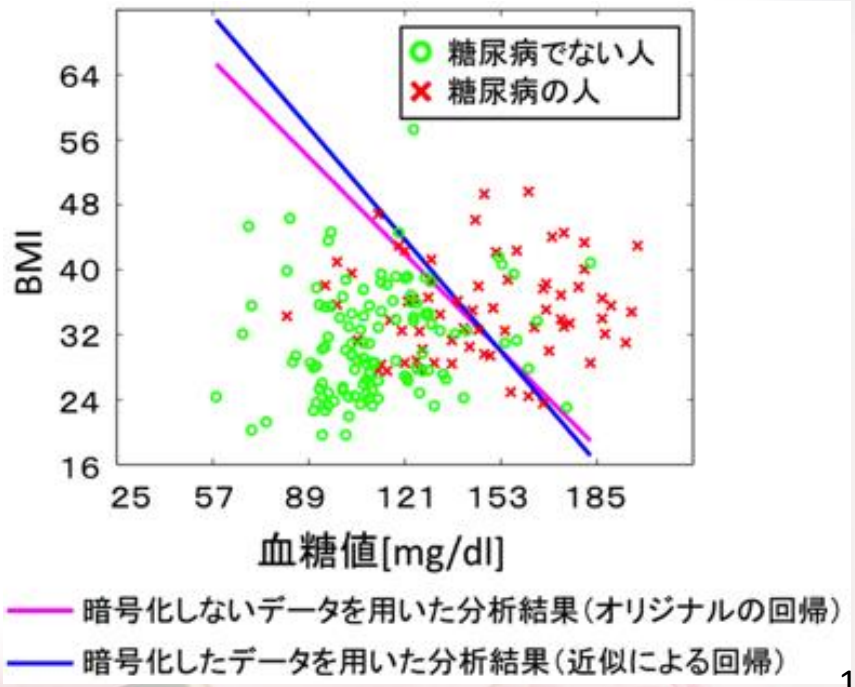




暗号化したままビッグデータ分類

- ビッグデータ解析で多用されているロジスティック回帰分析をデータを暗号化したまま計算可能に
- 暗号化された1億件のデータを30分以内で複数グループに分類できることをシミュレーションで確認
  - NICTプレスリリース「暗号化したままデータを分類できるビッグデータ向け解析技術を開発」(2016.1.14)





暗号化したまま医療データ解析

- 病気の罹患情報と個人の遺伝情報との統計的な関連性を個人の病気の有無を知ることなくχ^2検定で解析
- 4500名程度の規模で1分弱
- 解析対象外データが混在した場合でも高速検出（数十ミリ秒）
  - 2018.7.18 JST, 筑波大と共同プレスリリース

①医療データを暗号化

病気の罹患情報

	糖尿病	高血圧
A	あり	なし
B	あり	あり
C	:	:

病院

②医療データの暗号文を送付

遺伝情報を管理する検査機関

個人の遺伝情報

	rs001	rs002
A	あり	なし
B	あり	あり
C	:	:

③個々の医療データを知ることなく遺伝情報との関連性（統計値）の暗号文を計算

④統計値の暗号文を送付

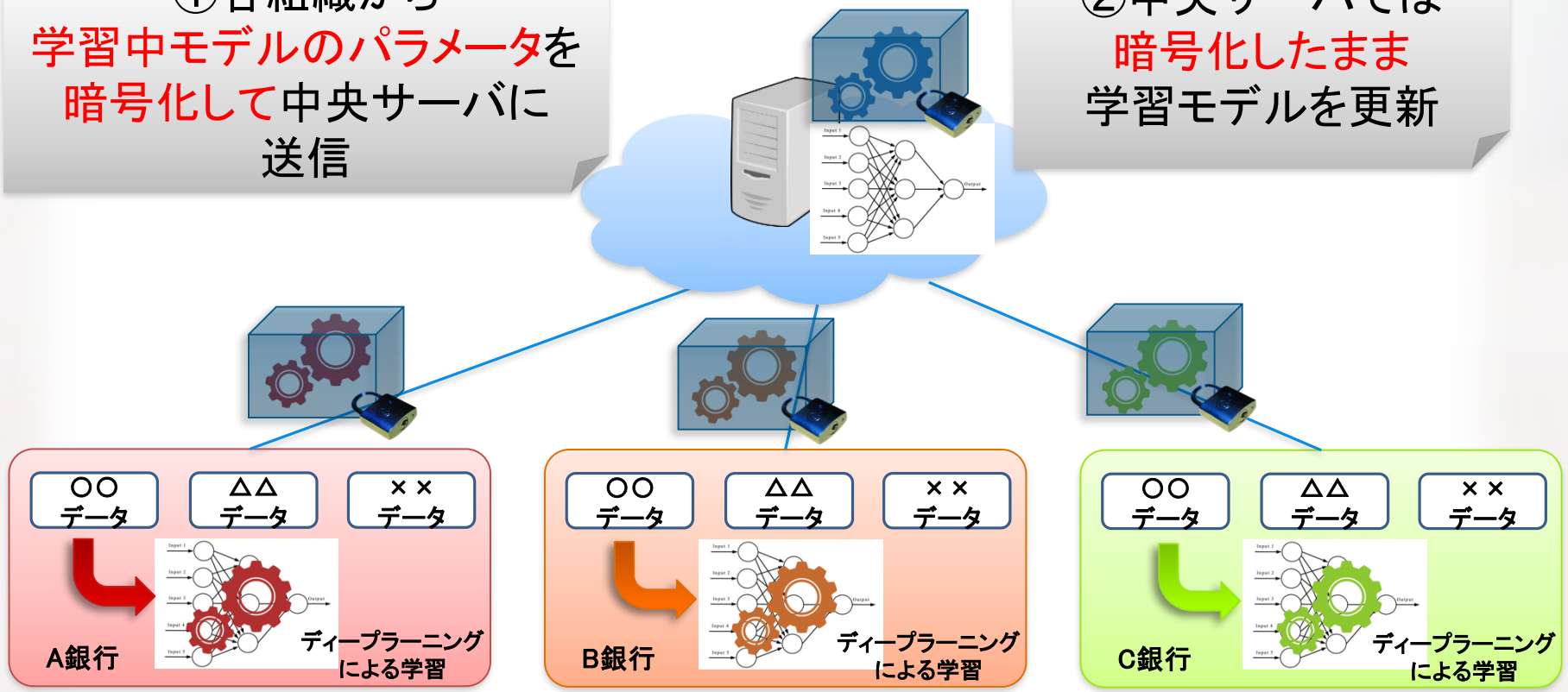
⑤個々の遺伝情報を知ることなく医療データとの関連性を得る

異なる病気の場合 警告11

DeepProtect:
外部にデータ開示することなく
複数組織で協調して深層学習

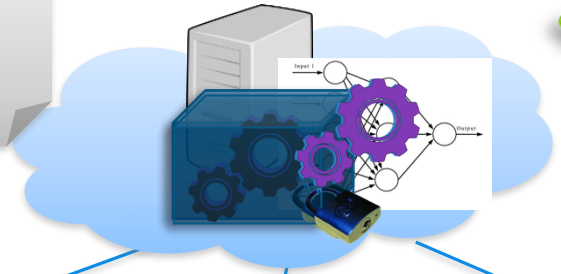
①各組織から
学習中モデルのパラメータを
暗号化して中央サーバに
送信

②中央サーバでは
暗号化したまま
学習モデルを更新

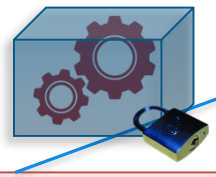


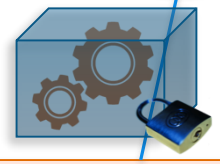
DeepProtect:
外部にデータ開示することなく
複数組織で協調して深層学習

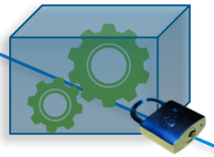
③各組織では
中央サーバで更新された
学習モデルをダウンロード、
精度の高い分析が可能に



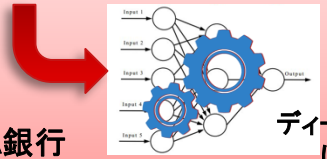
中央サーバにも
第三者にも
データ漏洩の
危険性なし！





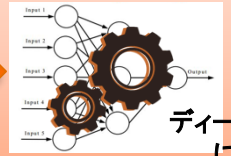


○○ データ　△△ データ　×× データ



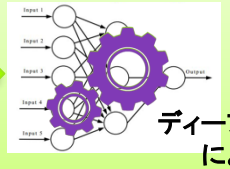
A銀行　ディープラーニング による学習

○○ データ　△△ データ　×× データ



B銀行　ディープラーニング による学習

○○ データ　△△ データ　×× データ



C銀行　ディープラーニング による学習

まとめ

- 新型コロナウイルス感染症拡大防止と“ニューノーマル”に向け**ICTとデータの活用**を

- **プライバシー**に配慮したデータの利活用
- Withコロナ期にプライバシーを保護し**長期間**いかにデータを利活用する技術を社会に実装し**「実践」**していくか。