NIIワークショップ
研究・教育のためのデータ連携ワークショップ（第1回） 講演
2009-04-22

データ連携セキュリティの課題

国立情報学研究所(NII)
コンテンツ科学研究系
越前 功

研究･教育データの連携

- データ連携により新たな研究開発の創生，情報サービスの提供が進む
  - 医療・保険・福祉データの連携
  - 「想・IMAGINE」：書誌DB，博物館・美術館文化財DBの連携
  - 「デジタル台風」：気象映像画像DB，台風経路・勢力データ，アメダス観測データ，ニュース記事DBの連携

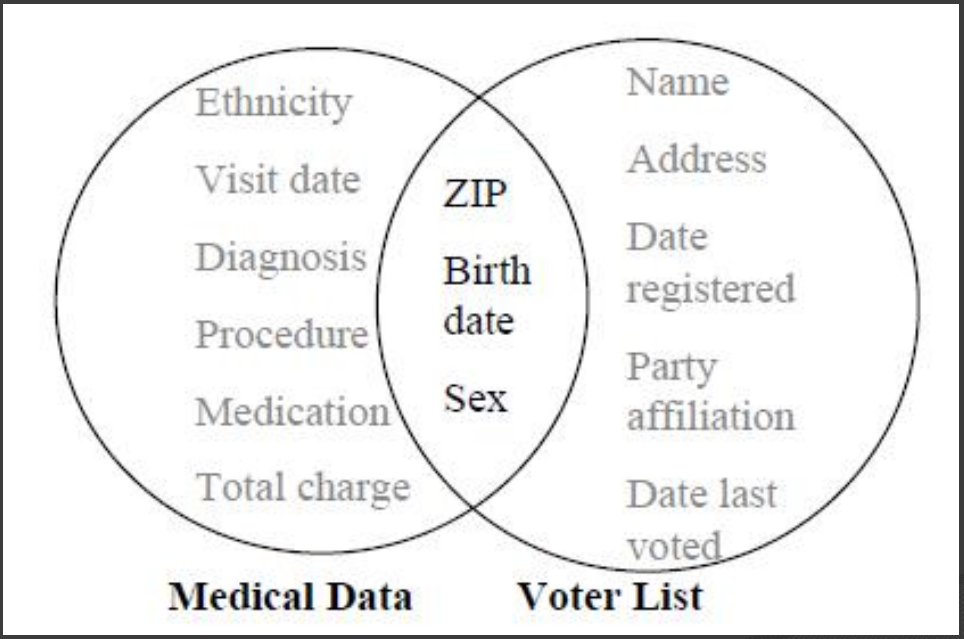
- 研究・教育データのセキュリティ対策
  - データ充実やデータ連携が進むと企業の業務データと同様なセキュリティ対策が必要
    - 匿名性確保，プライバシ保護，アクセス制御など
  - 大学，研究機関のセキュリティ意識の向上
    - 情報セキュリティポリシーの策定

研究･教育データのライフサイクルとセキュリティ課題

ライフサイクル	課題	対策
登録・収集	匿名性の確保	匿名ID，匿名署名（グループ署名）
活用（データ連携）	個人情報などの非公開データの推測防止（プライバシ保護）	推論制御，k-匿名性
保存	-データの改ざん防止 -非登録者の利用防止	WORM(Write Once Read Many)，アクセス制御，フロー制御
廃棄	データ復元防止	データ抹消，メディア廃棄

データ連携におけるプライバシ保護の必要性

- データ連携により個人情報が推測されてしまう
  - 医療データ，投票者情報の連携による個人情報の推定
  - 連携前の各データ管理者が適切に対策しても，連携により推定できてしまう恐れあり



出典：L. Sweeney, k-Anonymity: A Model for Protecting Privacy, Intl J on Uncertainty, Fuzziness and Knowledge-based Systems, vol. 10, no. 5, pp. 557-570, 2002

データ連携におけるプライバシ保護の必要性

- データ連携により個人情報が推測されてしまう
  - 医療データ，投票者情報の連携による個人情報の推定
  - 連結前の各データ管理者が適切に対策しても，連結により推定できてしまう恐れあり
- 対策：ポリシーの策定
  - データ連携毎に以下の見直し実施
    - サービスの目的：個別データの公開 or 統計データの公開
    - 非公開データ（不必要にデータを公開していないか？）
    - データ提供者の承諾
- 対策：データベースセキュリティ技術の適用
  - 推論制御 (inference controls)
  - k-匿名性 (k-anonymity)

推論制御と k-匿名性

- 推論制御 (inference controls)
  - 非公開情報が推定されないように公開情報の提供を制限する仕組み
    - 攻撃者：質問文を試行錯誤して，要素数 *S* が小さい公開情報を得ようとする（例：公開情報：地域別住民数，収入分布，質問文：年収1億以上の特定地域の住民数は？→当該地域の立派な住宅と関連付けてその住民の年収を推定）
    - 防御者：要素数 *S* が“*S* < *t* または *S* >= *N* - *t*” の場合，当該情報の提供を拒否

- k-匿名性 (k-anonymity)
  - 個人情報を格納したデータベースの各行において，当該行が持つ属性情報と同じ情報を持つ行が当該行を含めk個以上ある状態

犯罪歴	生年	性別	居住地
無	1971	女	152-8550
無	1971	女	152-8550
有	1973	男	101-8430
無	1974	女	215-0013
無	1971	男	101-8430
無	1974	男	215-0013
無	1972	男	101-8430

犯罪歴	生年	性別	居住地
無	1971	女	152-8550
無	1971	女	152-8550
不明	197-	男	101-8430
無	1974	不明	215-0013
不明	197-	男	101-8430
無	1974	不明	215-0013
不明	197-	男	101-8430

2-匿名性を満たすテーブル変換の例

さらなる課題

- データ価値とプライバシ保護の両立
  - トレードオフの関係
    - プライバシ保護重視：データの研究・教育価値を損なう
    - データ価値重視：データ提供者がデータを提供しにくくなる

- 集合知を用いた非公開情報推測攻撃への対策
  - 検索，Webマイニングを用いて教育・研究データから個人情報を推測される可能性がある
    - SNS，フィッシング検知では当該攻撃への対策が検討されている

参考文献

[1] 小松，プライバシ保護のためのアーキテクチャ，情報処理，vol. 48，no. 7, pp.737-743 , 2007

[2] D. Denning, Cryptography and Data Security, Addison-Wesley,1982

[3] 原嶋，鈴木，システムリスク管理におけるデータベース技術，東芝レビュー，vol.58, no.8, 39-42, 2003

[4] L. Sweeney, k-Anonymity: A Model for Protecting Privacy, Intl J on Uncertainty, Fuzziness and Knowledge-based Systems, vol. 10, no. 5, pp. 557-570, 2002

[5] 村本，上土井，若林，k-匿名性を利用したデータ一般化によるプライバシー保護，DEWS2007, A7-10, 2007

[6] 吉浦，片岡，中山，多様化するメディア環境に適応するヒューマンコミュニケーションセキュリティの構想,2007