

調査票情報の二次的利用の 利便性向上に関する意見

南 和宏
統計数理研究所

磁気媒体の提供の申請手続きを 短縮するための案

- 申請者に対し、審査基準を明確にする
 - 出力するデータの安全性
 - 最小権限の原則(Principle of least privilege)
- 事前チェック → 公表前の事後チェック
 - まだ作成していない分析結果の安全性審査は非常に困難
 - 審査担当者の事前分析が必要
- 変数単位の利用妥当性確認の簡素化

利用妥当性の簡素化 (ブラックボックス方式)

- 調査票情報に含まれる外観識別性の高い変数、機密性の高い変数のみを確認
- 残りの変数の利用は基本的に認める

申請手続きの期間短縮

- 審査作業のための人員が不足
 - 技術的な専門性を継続的に高められる専門職の育成、キャリアパスの整備

リモートアクセス方式

- 研究者の利便性向上、中間データの審査作業削減の観点からできるだけ早い移行が望まれる
- ただし、利用の絶対量が増えると予想されるので審査体制の拡充は不可欠
- 現状のオンサイト利用をそのままリモートアクセス方式に移行するのは不十分

リモートアクセスに向けた技術的課題

- 各調査票情報に共通なメタデータの整備
- ファイルベースの利用から統一的なデータベースへの移行
 - データベース理論に基づくスキーマ(メタデータ)の正規化
 - 冗長性の除去
 - アクセス権限の管理による利用制限
 - 現状はユーザーごとにデータを切り出してファイルとして提供
 - 共通ID(おそらくマイナンバー)による自由なテーブルの結合(リンケージ)の実現

データベースのスキーマの正規化

元の
スキーマR



正規化

変換された
スキーマR*

R* must:

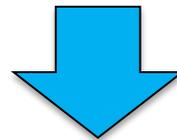
- 元のスキーマRの情報を保持
- 冗長性を最小化
- 変数間の依存性を保持
- 優れたシステム性能を提供

正規化の理論では「冗長性」の概念が定式化されている

正規化の例

レコードの
長さが可変

| <u>SSN</u> | Name | Phone Number |
|-------------|------|----------------------------------|
| 123-32-1099 | Fred | (201) 555-1234 (206) 572-4312 |
| 909-43-4444 | Joe | (908) 464-0028 (212) 555-4000 |



第1正規化

| <u>SSN</u> | Name | <u>Phone Number</u> |
|-------------|---------|---------------------|
| 123-32-1099 | Fred | (201) 555-1234 |
| 123-32-1099 | Fred | (206) 572-4312 |
| 909-43-4444 | Joe | (908) 464-0028 |
| 909-43-4444 | Joe | (212) 555-4000 |
| 234-56-7890 | Jocelyn | (212) 555-4000 |

| <u>SSN</u> | Name | <u>Phone Number</u> |
|-------------|---------|---------------------|
| 123-32-1099 | Fred | (201) 555-1234 |
| 123-32-1099 | Fred | (206) 572-4312 |
| 909-43-4444 | Joe | (908) 464-0028 |
| 909-43-4444 | Joe | (212) 555-4000 |
| 234-56-7890 | Jocelyn | (212) 555-4000 |

データの不一致
の原因

Boyce-Codd正規化

| <u>SSN</u> | Name |
|-------------|------|
| 123-32-1099 | Fred |
| 909-43-4444 | Joe |

| <u>SSN</u> | <u>Phone Number</u> |
|-------------|---------------------|
| 123-32-1099 | (201) 555-1234 |
| 123-32-1099 | (206) 572-4312 |
| 909-43-4444 | (908) 464-0028 |
| 909-43-4444 | (212) 555-4000 |