2010 통계자료의 개인정보보호 국제회의 참가 결과

(2010 Privacy in Statistical Databases: PSD)

2010. 10

통계청 STATISTICS KOREA

<차 례>

I 출장개요

1. 회의 개요

❑ 회의명 : 2010 통계자료의 개인정보보호 국제회의
(2010 Privacy in Statistical Databases: PSD)

❑ 기간 및 장소

○ 2010. 9. 22(수) ~ 9. 24(금), 그리스 코르푸

❑ 연혁 및 규모

○ 2004년부터 매 짝수 년 개최되어 이번 PSD2010은 제4차 회의임

◦ PSD2004(Privacy in Statistical Databases) : 2004. 6.9~11, 스페인 바르셀로나
◦ PSD2006 : 2006. 12.13~15, 이탈리아 로마
◦ PSD2008 : 2008. 9.24~26, 터키 이스탄불

○ 16개국 44개 기관의 52명 참가
- EU 회원국과 미국을 중심으로 한 주요 선진국 통계전문가

○ 비밀보호방법론과 관련된 국제회의는 크게 2개로, PSD (Privacy in Statistical Databases)가 매 짝수 년 개최되고 이와 더불어 UNECE에서 주관하는 Confidentiality Conference가 매 홀수 년 개최됨. UNECE/Confidentiality Conference가 국가기관의 데이터 비밀보호 사례에 관한 논문이 비교적 많은 반면 본 회의는 비밀보호 방법론에 대한 학술적인 연구에 관한 논문이 많음 편임

2. 출장자 : 연구기획실 김경미 주무관, 정보서비스팀 이필영 주무관

3. 회의주제 : 『자료의 비밀보호방법』

❑ 주요 세션 내용 (9.22~24)

<table>
<tr><th></th><th>9.22 (수)</th><th>9.23 (목)</th><th>9.24 (금)</th></tr>
<tr><td>9:00</td><td>등록/기조연설</td><td rowspan="2">마이크로데이터
비밀보호 방법론Ⅰ
(Microdata
Protection Ⅰ)</td><td rowspan="2">온라인 데이터베이스와
원격접근
(On-Line Data Bases
and Remote Access)</td></tr>
<tr><td>10:00</td><td>사생활 보호 규정
(Privacy-Preserving
Protocols)</td></tr>
<tr><td>11:00</td><td>휴식</td><td>휴식</td><td>휴식</td></tr>
<tr><td>11:25</td><td rowspan="2">집계표데이터
비밀보호 방법론Ⅰ
(Tabular Data
Protection Ⅰ)</td><td rowspan="2">마이크로데이터
비밀보호 방법론Ⅱ
(Microdata
Protection Ⅱ)</td><td rowspan="2">법적 이슈
(Legal Issues)</td></tr>
<tr><td>12:00</td></tr>
<tr><td>12:40</td><td>점심</td><td>점심</td><td rowspan="8"></td></tr>
<tr><td>14:45</td><td rowspan="3">집계표데이터
비밀보호 방법론Ⅱ
(Tabular Data
Protection Ⅱ)</td><td rowspan="7">차별적 사생활 보호
(Differential Privacy)</td></tr>
<tr><td>15:00</td></tr>
<tr><td>15:30</td></tr>
<tr><td>16:00</td><td>휴식</td></tr>
<tr><td>16:30</td><td rowspan="3">합성데이터 방법론
(Synthetic Data)</td></tr>
<tr><td>17:00</td></tr>
<tr><td>17:30</td></tr>
</table>

II 주요 회의내용

1. 사생활보호 규정(Privacy-Preserving Protocols)

❑ 개인의 비밀 보호를 위한 협동 사생활 방법과 레코드 연결 기법에서 개인정보 보호 방법에 관한 주제 발표

❑ 발표 논문

❍ Coprivacy: towards a Theory of Sustainable Privacy, Josep Domingo-Ferrer (UNESCO chair in Data Privacy)

❍ Privacy-Preserving Record Linkage, Rob Hall and Stephen E. Fienberg (미국 - 카네기 멜런 대학교)

❍ Techniques for Content Subscription Anonymity with Distributed Brokers, Sasu Tarkoma and Christian Prehofer

❑ 내용 요약

❍ 사생활 보호를 위하여 협동사생활(coprivacy) 또는 협동조직 사생활(co-operative privacy)의 내용이 주목되어짐. 개인의 사생활을 보호하는 최선의 방법이 다른 사람의 사생활을 보호하는 것에 도움이 된다면 협동사생활임. 합동사생활은 i) 도와주는 개인의 효용은 사생활, 비밀, 소득기능을 포함함 ii) 혼합전략과 혼합내쉬균형(Nash-euilibria)은 몇몇 제한을 허용하는 것을 혼합 공동사생활임 iii) 내쉬균형(Nash-euilibria)과 상관되는 협동사생활은 상관균형에 의함. 협동사생활은 개인 대 개인(P2P) 규약에 적용됨.

❍ 레코드연결은 하나의 파일 응답과 다른 파일의 레코드가 같은 사람을 설명할 수 있는가를 결정하는 것으로 최근 데이터 통합, 통합데이터 계산 시 중간단계, 통합된 데이터를 분석을 위한 공공이용 파일생성을 목적으로 레코드 연결을 한다. 따라서 레코드 연결 시 공동변수에 대한 암호변수, 노이즈에 의해 변형된 자료 생성 등으로 사생활이 더 보호되어짐.

2. 집계표 데이터 비밀보호(Tabular Data Protection)

❑ 주요 발표내용

1) 실현 불가능한 혹은 구현 불가능한 RCTA 사례를 분석하고 해결하기 위한 툴 (A Tool for Analyzing and Fixing Infeasible RCTA Instances)

Jordi Castro and Jose A. Gonzalez

(스페인 바르셀로나 - 카탈로니아 대학교)

❍ RCTA Instance : 카탈로니아(catalonia)정부 과학혁신성(ministry of science and innovation)에서 제안된 방법

- 집계표데이터의 비밀보호기법 중 최근에 많이 연구되어지는 방법론인 일반적인 CTA (Controlled Tabular Adjustment) 방법을 응용한 Restricted CTA방법

❍ 최소거리 CTA방법과 RCTA방법은 최근 집계표데이터의 비밀보호를 위한 변조적인(perturbative) 방법론임.

❍ 비밀보호된 집계표가 주어졌을 때, RCTA방법의 목적은 민감한 셀을 위한 보호 수준을 결정하는 closest table을 찾는 것임. 이 closest table은 남은 셀들에 미미한 조정함을 통해 얻음. 가능하면 그들의 실제값의 subset(대게 전체 셀)을 포함하지 않으면서, 만약 비밀보호 수준이 크거나 또는 셀의 편차의 바운더리가 타이트하거나 또는 너무 많은 셀을 보존해야 한다면 결과적인 혼합된 정수형 선형 문제(mixed integer linear problem)가 실현 혹은 구현 불가능한 것으로 나오게 됨.

❍ 이 방법은 구현 불가능한 사례를 분석하기 위한 발전된 툴을 설명하고 있음. 이 툴은 일반적인 유연한 프로그래밍 접근법에 기초하며 이는 여유있는 제약을 얻는 인위적인 문제와 남은 유연한 변수를 추가한 범위를 고려함.

2) 셀 감추기 방법에서의 가지-절단 방법과 절단-가지 방법의 비교 (Branch-and-Cut versus Cut-and-Branch Algorithms for Cell Suppression)

Juan-Jose Salazar-Gonzalez

(스페인 - 라 라구나 대학교)

❍ 이들 방법의 주요한 쟁점은 $\sum_{i \in I \setminus P} w_i x_i$ 을 어떻게 최소화하느냐하는 문제임.

- master problem :

$$\text{minimize} \sum_{i \in I \setminus P} w_i x_i \ , x_i \in \{0, 1\} \ , {}^{\forall} i \in I/P$$

❍ 절단-가지 방법(Branch-and-Cut: B&C)과 가지-절단 방법(Cut-and-Branch: C&B)의 비교

- 절단-가지(C&B)방법이 가지-절단(B&C)방법에 비해 이해하고 실행하기에 더 쉬움.
- 이 논문의 예로는 C&B방법이 B&C방법보다 더 경쟁적이고 완전성이 높게 나타남.
- 실제 표를 보호하는 방법으로 FOSS(Free Open Source Software)를 이용하여 설명함.

3) 센서스 집계표를 비밀보호하기 위한 데이터 교환 (Data Swapping for Protecting Census Tables)

Natalie Shlomo, Caroline Tudor and Paul Groom

(영국 - 사우스햄튼 대학교)

❍ 데이터교환(Data Swapping)방법은 크게 표적데이터 교환 전략(Targeted Data Swap Strategy)과 임의데이터 교환 전략(Random Data Swap Strategy)으로 나누어짐.

❍ 이 논문은 표적데이터 교환 전략을 제안함

- 이 방법은 데이터 교환 시에 위험이 높은 가구에 확률적으로 선택비율을 높임.
- 평등한 가구의 경우 지리적인 계층적 측면에서 위험의 수준에 따라 결정함.
- 노출위험과 데이터의 효용성 관점에서 이 방법을 추천함.

4) 센서스 빈도표의 소수 셀 제거: 변환 확률의 고려사항 (Eliminating Small Cells From Census Counts Tables: Some Considerations on Transition Probabilities)

Sarah Giessing and Jorg Hohne

(독일 - 독일통계청, 베를린-브란덴부르크 통계연구원)

- ❍ 2011년 독일센서스를 준비하는 과정에 센서스 빈도표에 적용할 변조방법을 비교연구하기 위해 시작됨.
- ❍ 마이크로데이터로부터 집계된 빈도가 소수의 셀일 때(예를 들어 빈도가 1 또는 2일 때)의 경우에 대해,
- ❍ 본 논문에서 제안하는 사전집계-임의변조방법(post-tabular random perturbation)에 기초한 마이크로데이터 키변수 변화량 맥락에서의 변환행렬로 집계된 방법의 결과로부터 측정된 변환 확률을 경험적으로 비교함
- ❍ SAFE라는 소프트웨어를 이용함
  - 이 소프트웨어는 베를린-브란덴부르크 통계연구원에서 개발되어 수년간 사용되어오고 있음.
  - 셀 빈도의 변조값을 산출하는 알고리즘이 포함되어 있음.

5) τ-Argus를 이용하여 연결된 SBS 집계표의 데이터셋을 다루는 세가지 방법 (Three Ways to Deal with a Set of Linked SBS Tables Using τ-Argus)

Peter-Paul de Wolf and Anco Hundepool

(네덜란드 - 네덜란드 통계청)

❍ SBS : Structural Business Statistics (구조적 경제 통계)

❍ 구조적 경제 통계(SBS) 집계표는 각각의 통계가 연결된 형태로 많이 존재함.

❍ 이러한 연결된 집계표의 정보노출을 제한하는 새로운 비밀보호기법을 소개하고 이를 다른 두 방법과 비교함.

- 비밀보호기법 적용에 τ-Argus 프로그램을 이용함.

6) SAS 프로그램 내의 매크로를 이용한 τ-Argus 모듈 추가 사례 (Techniques for Using τ-Argus Modular on Sets of Linked Tables - SAS implementation)

Katrin Schmidt and Sarah Giessing

(독일 - 독일통계청)

❍ 데이터비밀보호기법 적용에 사용되는 τ-Argus 소프트웨어를 SAS 프로그램의 매크로를 이용하여 구연하는 방법을 소개함.

3. 합성데이터 방법론(Synthetic Data)

❑ 합성데이터의 비밀노출 감소사례, 소지역 통계 및 모집단 자료에 대한 합성데이터 생성에 대한 주제 발표

❑ 발표 논문

❍ Using Support vector Machines for Generating Synthetic Datasets, Jorg Drechsler (독일 - 노동연구원)

❍ Synthetic Data for Small Area Estimation,

Joseph W. S. and Trivellore E. R. (미국 - 미시건 대학교)

❍ Disclosure Risk of Synthetic Population Data with Application to EU-SILC 183, Matthias Templ and Andreas Alfons (오스트리아 - 비엔나 과학 대학교, 통계청)

❍ Synthetic Longitudinal Datasets. An Application of Sequential Regression and CART Models to German Business Data, Hans-Peter Hafner and Rainer Lenz

❑ 내용 요약

❍ 합성데이터는 데이터 제공에 있어서 혁신적인 접근임. 범주형 데이터의 합성데이터를 생성하기 위해 지지벡터방식(Support Vector Machines)이 매우 유용하며, 독일의 사업체 조사를 실증연구한 사례에서 표준모수모델링(standard parametric modeling)보다 데이터노출이 감소되었음.

❍ 지역수준의 정책결정을 위해 소지역의 마이크로데이터 접근을 요구함. 통계기관은 개인정보 및 비밀보호 때문에 공공이용데이터에 자세한 지역정보를 막고, 제한된 데이터를 마이크로데이터 이용센터에서 이용하는 불편한 방법으로 제공하였다. 이러한 대안으로 소지역 추정을 위해 상세한 지역정보를 포함하는 완전합성(fully-synthetic) 공공이용 마이크로데이터를 제공한다. 사후예측분포로부터 합성데이터를 만들기 위해 베이지안 계층모형을 이용하였음.

❍ 모집단 데이터는 일반적으로 접근할 수 없음. 근접하고 현실적인 모집단 데이터를 만드는 방법과 이러한 합성모집단에서 비밀보호를 5가지 최악의 시나리오를 제시하고, 이러한 경우에도 합성모집단은 비밀보호 되는 우수한 데이터임을 제시함

4. 마이크로데이터 비밀보호(Microdata Protection)

❑ 주요 발표내용

1) 점진적 알고리즘을 이용한 PRAM 최적화 (PRAM Optimization Using an Evolutionary Algorithm)

Jordi Mares and Vicenc Torra

(스페인 카탈로니아 - IIIA[1], CSIC[2])

[1] IIIA : Institute d'Investigacio en Intelligencia Artifical

[2] CSIC : Consejo Superior de Investigaciones Cientificas

❍ PRAM : Post Randomization Method (사후 확률화 방법)

- PRAM은 1997년에 처음 소개된 이래로 적어도 통계적 범주형 데이터 비밀보호에서는 여전히 적용되고 있는 방법임.
- 이는 훌륭한 비밀보호(good protection)를 위해서 좋은 변환 행렬(good transition matrix)을 얻어야하는 어려움 때문임.

❍ 이 논문에서는 최적(best) 행렬을 찾기 위해 정보의 손실과 노출위험을 동시에 고려하면서 점진적 알고리즘(evolutionary algorithm)을 이용해 어떻게 더 나은 비밀보호를 할 것인지에 대해 논함.

Evolutionary Algorithm to Enhance PRAM Matrices

```
Input: P(0) = X initial population
Output: P(t) = X′ final population
t ⇐ 0
evaluate(P(0))
while stopping(P(t)) ≠ true; do
 alter ⇐ randomly choose between mutation and cross
 if alter by mutation then
   X′ ⇐ mutation(X)
 else
   X′ ⇐ cross(X)
 end if
 evaluate(X, X′)
 t ⇐ t + 1
end while
return P(t)
```

❍ 이 방법의 적용을 경험적으로 평가하기 위해 100개 레코드의 실제 자료를 이용한 사례를 제공하고 있음.

2) 승법 잡음 법칙 (Multiplicative Noise Protocols)

Anna Oganian

(미국 - 조지아 남부 대학교)

❍ 통계생산기관들은 조사응답자가 제공하는 개인정보를 보호해야하는 의무와 연구활동을 위한 유용한 데이터의 제공 사이에서 갈등을 가짐. 이 때문에 마이크로데이터를 제공할 때 데이터에 통계적정보노출제한방법을 적용하는데 이 방법에는 잡음추가, 데이터 교환, 마이크로애그리게이션과 같은 방법을 사용됨. 본 논문은 이 중 여러 가지 승법잡음방법에 대해 설명함.

3) 측정오차와 통계적 정보노출제한 (Measurement Error and Statistical Disclosure Control)

Natalie Shlomo

(영국 - 사우스햄튼 대학교 사우스햄튼 통계과학연구원)

❍ 영국의 통계작성기관은 정보노출제한(Statistical Disclosure Control: SDC)방법을 적용한 후에 연구자들에게 마이크로데이터를 제공함.

❍ 잡음추가방법을 많이 이용하는데 잡음추가는 정보노출제한(SDC)방법 중 변조적인(perturbative) 방법론을 이용함. 이 방법은 연속형 변수에 독립적인 임의의 잡음을 추가하거나 확률 메커니즘에 따라 범주형 변수의 값을 오분류하는 방법임.

❍ 잡음을 추가함으로써 나타나는 이 오차는 고의로 발생시키는 것이므로 이 변조의 모수는 알려져 있고 측정오차 모형을 통해 연구자가 적합한 통계적 추론을 할 수 있음.

❍ 이러한 잡음을 추가하는 방법을 쓰는 반면, 통계작성기관이 변조모수를 제공하지 않는 경우가 대부분임.

❍ 이러한 문제를 해결하기 위해 본 논문에서는 변조된 데이터셋에 유효한 추론을 선험적으로 보장하는 수정된 정보노출제한방법을 제안하고 있음.

4) 데이터 환경분석과 핵심변수 맵핑 시스템 (Data Environment Analysis and the Key Variable Mapping System)

Mark Elliot, Susan Lomax, Elaine Mackey and Kingsley Purdam (영국 - 맨체스터대학교 조사연구센터)

❍ KVMS : Key Variable Mapping System(핵심변수 맵핑 시스템)

❍ 통계적 노출 위험 측정은 어떤 제안된 데이터셋이 공표되는 데이터 환경에 대한 분석과 함께 진행되어야 한다는 것이 최근 일반적인 추세임.

❍ 어떠한 데이터이냐에 따라 침입자나 공격시나리오의 환경이 다르기 때문에 이러한 것들이 고려되어야 한다는 것임.

❍ 1999년 엘리엇과 데일은 이러한 시나리오 분석을 위한 원칙의 일반적인 집합을 설정하였으며 이때의 결과물은 핵심변수의 집합임

❍ 이 논문은 경험에 근거한 데이터 환경 분석 방법에 대해 설명함. 이 방법은 핵심변수의 목록을 만들기 위해 디자인 된, 이전에 가능한 것보다 훨씬 더 정도 좋은 사양을 가지는, 핵심변수 맵핑시스템(KVMS)의 프로토타입 툴로 작동됨.

5. 온라인 데이터베이스와 원격접근(On-Line Databases and Remote Access)

❑ 온라인 상에서 자료의 제공 시 비밀노출 관리방법과 미국 및 유럽의 원격접근법에 대한 주제 발표

❑ 발표 논문

- ❍ Remote Data Access and the Risk of Disclosure from Linear Regression: An Empirical Study,

  Philipp Bleninger, Jorg Drechsler and Gerd Ronning

  (독일 - 노동연구원, 튀빙겐대학교)
- ❍ The Microdata Analysis System at the U.S. Census Bureau,

  Jason Lusero and Laura Zayatz (미국 - 센서스국 통계연구과)
- ❍ Establishing an Infrastructure for Remote Access to Microdata at Eurostat, Wolf Heinrich Reuter and Jean-Marc Museux (오스트리아 - 비엔나 경제 대학교, 룩셈부르크 - Eurostat 방법론 연구 조직)
- ❍ Improvement of data access. The long way to remote data access in Germany, Maurice Brandt and Markus Zwick
- ❍ Security of statitical databases as an element of an enterprise security architecture, Lukasz Slezak and Jaroslaw Butanowicz

❑ 내용 요약

❍ 연구자가 쉽게 데이터 접근을 위한 노력으로 원격데이터접근은 데이터 제공 전에 데이터를 바꾸거나 데이터 아카이브 또는 데이터 센터 이용에 대한 대안임. 실제로 마이크로데이터를 보지 않고 쿼리를 제공하는 원격데이터접근 또는 원격분석 서비스는 마이크로데이터를 이용자가 직접 이용하지 않을지라도 응답자의 민감정보 누출은 여전히 가능함. 원격데이터접근 제공자는 시스템에서 허용하는 쿼리를 주의 깊게 고려해야 함.

❍ 미국 센서스국은 응답자의 비밀을 누설하지 않고 데이터를 제공해야하는 의무를 가지고 이용자가 마이크로데이터를 보거나 다운로드 하지 않고 센서스자료를 다양한 통계분석(회귀분석, 교차분석, 상관관계 등)하도록 마이크로데이터 분석 시스템(MAS: Microdata Analysis System)을 사용함.

❍ Eurostat는 마이크로데이터 수요에 대한 만족화를 위해 이용자에게 (1)터미널 서버 (2)원거리 네트워크 (3)작업반출 시스템을 기반으로 원격접근을 제공하고 있음. IT구조에서 워크스테이션, 이용자 매니지먼트와 신뢰, 파일시스템, 노출관리의 Eurostat 도전 및 노력이 필요함.

6. 법적 문제(Legal Issue)

❑ 유럽의 통계시스템에서 마이크로데이터 파일의 통계적 누출에 법적, 방법론적인 현안문제에 대한 주제 발표

❑ 발표 논문

❍ Strategies to Achieve SDC Harmonisation at European Level: Multiple Countries, Multiple files, Multiple Surveys, Daniela Ichim and Luisa Franconi (이탈리아 - 통계연구원)

❑ 내용 요약

❍ 마이크로데이터 파일의 통계적 누출에 대해 방법론과 조직적부분에서 조화를 제안함. 방법론적(Perez-Duarte, 2009)으로는 익명화 과정에서 입력물과 결과물의 최소한의 요구이며 조직적 부분에서 유럽통계시스템의 법률을 공유하는 것임. 방법과 절차(조직) 체계에서 동일 통계조사, 동일 데이터셋, 다중파일 제공에 관하여 유럽 통계에 새로운 규정으로 마이크로데이터에 접근해야 함.

Ⅲ 시사점

❑ 자료의 비밀보호방법론은 최근 크게 마이크로데이터* 비밀보호방법론, 집계표데이터** 비밀보호방법론, 합성데이터*** 방법론으로 나누어짐

* 마이크로데이터(microdata) : 원시자료(raw data)에 입력오류나 조사오류를 수정한 자료로서 통계표 작성 등 데이터 가공의 바탕이 되는 자료

** 집계표데이터(tabular data) : 마이크로데이터를 임의의 기준에 따라 집계한 자료, 매크로데이터(macrodata)와 같은 의미이며 최근에는 이 용어보다 집계표데이터(tabular data)라고 많이 쓰이고 있음

*** 합성데이터(synthetic data) : 비밀보호 목적 등을 위해 합성한 데이터

❑ 미국 및 유럽의 국가들은 1980년대부터 비밀보호방법론에 대한 많은 연구가 진행되어 현재는 이를 바탕으로 각 방법론에 대한 더 심화되고 응용된 연구가 활발히 진행되고 있음

❍ 마이크로데이터 비밀보호방법론

- 기존의 확률화방법(randomization method), 잡음추가방법(addictive or multiplicative noise method) 등이 각 나라의 자료의 특성에 맞게 응용하거나 심화된 기법들이 소개됨
- 데이터의 특성에 따른 침입자(intruder)나 공격시나리오를 고려한 데이터 환경 분석(data environment analysis)에 관한 연구도 진행되고 있음

❍ 집계표데이터 비밀보호방법론

- 기존의 데이터교환방법(data swapping), 셀감추기방법(cell suppression), CTA(controlled tabular adjustment)방법 등에 대한 심화연구가 활발함

- 각 방법론들을 프로그램으로 구현하는 방법에 대한 연구도 활발(별도의 프로그램을 만들거나, 기존의 프로그램을 일반적 통계패키지 SAS의 매크로를 이용해 구현해보는 등)
- 비밀보호 시 빈도표(counts tables)에서 많은 문제를 안고 있는 소수셀(small cells : 빈도가 1 또는 2, 혹은 특정 수 이하)에 관한 연구도 진행되고 있음

❑ 비밀보호에 대한 연구는 미국과 유럽을 중심으로 다양한 방법론이 개발되어 수 십년에 거쳐 이에 대한 꾸준한 심화연구가 이루어지고 있음

❑ 사생활보호의 개념이 점점 강화되고 있는 시점에서 우리청에서도 비밀보호방법론에 관한 체계적이고 지속적인 연구가 필요함

❑ 자료의 제공방법이 다양해짐에 따라 온라인접근 및 원격접근 및 여러 가지 통계자료에 관한 연구도 활발히 진행되고 있으며 이에 관한 관심도 필요함

❍ 원격접근에 의한 자료제공 시 비밀보호 및 쿼리 관리
- 원격접근으로 연구자들이 사용한 쿼리를 정확히 검토하고 분석결과 집계표에서 비밀정보사항이 누출되지 않는 철저한 관리가 필요

❍ 소지역통계자료 제공에 따른 비밀보호기법 적용
- 소지역통계자료는 정보의 노출위험이 더욱 크므로 합성데이터 기법 등 이에 따른 비밀보호기법 연구 필요

❍ 레코드 연결에 따른 자료비밀보호

- 최근 서로 다른 조사 데이터를 결합하거나 혹은 행정자료와 연계하는 등의 데이터통합분석이 활발함에 따라 레코드 연결데이터 등의 통합자료에 대한 자료비밀보호기법 연구 필요