

Les appariements sécurisés

Catherine Quantin et Benoît Riandey – Lundi 17 novembre de 14h à 17h30

Les appariements de fichiers permettent de traiter conjointement des informations relatives au même individu, mais localisées dans plusieurs sources ou issues de vagues de collecte successives. En matière d'enquête, il s'agit principalement de compléter le questionnaire à l'aide d'informations administratives issues de la base de sondage ou de tout autre fichier administratif. Il peut au contraire s'agir de rassembler des informations de divers fichiers administratifs comme des chapitres différents d'un même questionnaire d'enquête. C'est alors un substitut économique à une enquête.

Il s'agira d'appariements sécurisés quand l'opération sera menée sans recours à un identifiant nominatif, soit du fait de son absence, soit par précaution ou obligation de sécurité. La technique la plus simple consiste à crypter de manière irréversible l'identifiant. On parlera alors de hachage et d'un appariement déterministe. Le hachage, avec la même clé secrète, fournira toujours le même identifiant anonyme, base de l'appariement des données d'individus non dévoilés.

On peut aussi recourir à plusieurs caractéristiques dont le croisement isole généralement un individu, mais avec une certaine incertitude. On parlera alors d'appariement probabiliste. Si les caractéristiques ont fait l'objet d'un hachage, il s'agit d'un appariement sécurisé probabiliste. Cette technique permet aussi d'achever un appariement déterministe quand l'identifiant présente des erreurs entravant le bon appariement de certaines unités.

Ces techniques ont d'abord été utilisées en épidémiologie et en économie médicale pour protéger le secret médical, en particulier pour masquer le numéro de sécurité sociale. Son recours en statistique sociale serait d'une très grande fécondité pour exploiter conjointement et anonymement des fichiers sociaux munis du même identifiant, mais transmis anonymisés par obligation juridique ou déontologique de confidentialité.

La présentation des méthodes permettra de citer de nombreuses applications et d'évoquer les potentialités à développer dans de nouveaux domaines statistiques.

Bibliographie

- Fellegi, I. P. and A. B. Sunter. "A theory for record linkage." *Journal of the American Statistical Association* 1969;64(328): 1183-1210.
- Fournel I., Schwarzingler M., Benzenine E., Binquet C., Riandey B., Hill C., Quantin C. Déterminer le statut vital des patients domiciliés en France métropolitaine et dans les départements d'Outre Mer en chaînant données hospitalières et données de mortalité de l'INSEE anonymisées - Conférence universitaire des démographes (CUDEP), Bordeaux, 21 au 24 mai 2007
- Fournel I, Schwarzingler M, Binquet C, Benzenine E, Hill C, Quantin C. Contribution of record linkage to vital status determination in cancer patients. *Stud Health Technol Inform.* 2009;150:91-5. PubMed PMID: 19745273.

- Gensbittel MH., Riandey B., Quantin C. Appariements sécurisés : statisticiens, ayez de l'audace ! *Courrier des statistiques* n° 121-122, mai-décembre 2007 :49-58.
- Goy A., « L'appariement sécurisé de fichiers d'étudiants grâce au hachage des identifiants », n° 113-114, p.23-32 mars-juin2005et vol 146 n°3, 2005.
- Guillerme M. L'appariement de fichiers pour le suivi de l'étudiant. *Mémoire professionnel, Formation Continue Diplômante des Attachés*, janvier 2009.
- Jaro M. Advances in record linkage methodology as applied to matching the 1985 census of Tampa, Florida. *Journal of the American Statistical Association*, 1989, 84, 414-420.
- Picard N, Riandey B, Solaz A, Quantin C. Les besoins des démo-économistes en matière d'appariements sécurisés. *Courrier des Statistiques* n°129, juin 2010.
- Quantin C, Allaert FA, Avillach P, Fassa M, Riandey B, Trouessin G, Cohen O. Building application-related patient identifiers: what solution for a European country? *International Journal of Telemedicine and Application* 2008, Article ID 678302, 1-5.
- Quantin C., Allaert F., Avillach P., Riandey B., Fieschi M., Fassa M, Cohen O. Proposal of a french health identification number interoperable at the European level. *Medinfo*, 2007, 12 : 503-7.
- Quantin C, Allaert FA, Avillach P, Riandey B, Fieschi M, Fassa M, Cohen O. Proposal of a french health identification number interoperable at the European level. *Risques et Qualité*, 2008;5:62.
- Quantin C, Allaert FA, Fassa M, Riandey B, Avillach P., Cohen O. How to manage a secure direct access of European patients to their computerised medical record and personal medical record. *Stud health Technol Inform* 2007;127:246-55.
- Quantin C, Bouzelat H, Allaert FA, Benhamiche AM, Faivre J, Dusserre L. Automatic record hash coding and linkage for epidemiological follow-up data confidentiality. *Methods Inf Med* 1998;37:271-7.
- Quantin C, Cohen O, Riandey B, Allaert FA. The French proposal for a health identification number. *Stud Health Technol Inform* 2006; 124:201-6.
- Quantin C, Cohen O, Riandey B, Allaert FA. Unique patient concept: a key choice for European epidemiology. *International Journal of Medical Informatics*, 2007;76:419-426.
- Quantin C, Fassa M, Coatrieux G, Riandey B, Trouessin G, Allaert FA. Chaînage de bases de données anonymisées pour les études épidémiologiques multicentriques nationales et internationales : proposition d'un algorithme cryptographique. *Revue d'Epidémiologie et de Santé Publique*, 2009 ;57 :33
- Quantin C, Fassa M, Coatrieux G, Trouessin G, Allaert FA. Combining hashing and enciphering algorithms for epidemiological analysis of gathered data. *Methods Inf Med*. 2008;47(5):454-8.
- Quantin C, Gouyon B, Allaert FA, Dusserre L, Cohen O. Méthodologie pour le chaînage de données sensibles tout en respectant l'anonymat : application au suivi des informations médicales. *Courrier Stat* 2005;113- 114:15-26 [*Journal de la SFdS* 2005;146:15-26].

Ateliers de formation – 17 et 21 novembre 2014

8ème colloque francophone sur les sondages - Dijon



- Quantin C, Gouyon B, Avillach P, Ferdynus C, Sagot P, GouyonN JB. Using discharge abstracts to evaluate a regional perinatal network: assessment of the linkage procedure of anonymous data. *International Journal of Telemedicine and Applications*, 2009: ID 181847.
- Quantin C, Jaquet-Chiffelle DO, Coatrieux G, Benzenine E, Allaert FA. Medical record search engines, using pseudonymised patient identity: an alternative to centralised medical records. *International Journal of Medical Informatics*, 2011 ;80 :e6-e11.
- Quantin C, Jaquet-Chiffelle DO, Coatrieux G, Benzenine E, Auverlot B, Allaert FA. Medical record : systematic centralization versus secure on demand aggregation. *BMC Medical Informatics and Decision Making*, 2011; 11 :18.
- Trouessin G, Allaert FA. FOIN: a nominative information occultation function. *Stud Health Technol Inform* 1997; 43:196–200.