

Une nouvelle méthodologie pour l'anonymisation des entrepôts de données spatiales : application aux données de biodiversité dans le contexte agricole

Loris Croce^{*,**,***} Laetitia Lemièr^{*,**,***} Sandro Bimonte^{*}, François Pinet^{*}

* Université Clermont, IRSTEA, TSCF, Aubière, France

prenom.nom@irstea.fr,

<https://www.irstea.fr/>

** Université Clermont Auvergne, Aubière, France

prenom.nom@etu.uca.fr,

<https://www.uca.fr/>

*** Ces auteurs ont contribué de manière égale.

Résumé. Dans cet article, nous nous intéresserons au problème de l'anonymisation de données agricoles géo-référencées. C'est un sujet qui n'est que peu ou pas abordé dans la littérature mais pourtant pertinent puisque le monde agricole est une source de données très importante. En effet, ces données peuvent être utilisées pour mesurer la biodiversité. L'objectif est de rendre accessible des données agricoles à des fins de recherches sans briser l'anonymat des personnes participant à l'étude, car ces données peuvent d'avérer sensibles. Nous tentons de répondre à cela via une technique spécifique d'agrégation.

1 Introduction

Les entrepôts de données spatiales (EDS) et les systèmes OLAP spatial (SOLAP) permettent l'analyse en ligne de grandes volumes de données géo-référencées (Malinowski et Zimányi, 2008). Les données dans les EDS sont stockées selon le modèle *spatio-multidimensionnel* qui définit les concepts de dimension spatiale (dimension qui présente des valeurs géométriques) et mesure spatiale (valeur géométrique). Aujourd'hui, de plus en plus de données sont disponibles via les nouveaux systèmes d'acquisition (capteurs, images satellites, etc.), les réseaux sociaux, les données open data et les données volontaires (i.e. données des observatoires). Dans ce contexte, la mise en place des EDS peut se faire classiquement en utilisant des données internes aux organisations/entreprises, mais aussi en utilisant ces nouvelles sources de données (Ravat et al., 2016). De nombreux systèmes d'acquisition sont installés sur des terrains agricoles ce qui fait du monde agricole une source de données importantes. L'analyse de ces données serait très instructifs. Toutefois, cela soulève des nouvelles problématiques liées aux aspects de confidentialité. En effet, pour pouvoir être utilisées, ces données doivent être anonymisées car elles peuvent ne pas appartenir aux organisations/entreprises et il faut pouvoir garantir la confidentialité des informations. C'est pourquoi il est primordial d'établir une chaîne de confiance avec les personnes volontaires sur le respect de leur confidentialité. L'anonymisation