

Prise de décision à partir de données environnementales imparfaites

André Miralles¹, Franck Ravat² et Thérèse Libourel³

¹ Irstea – UMR Tetis - 34093 Montpellier Cedex 05 - andre.miralles@teledetection.fr

² Université Toulouse I Capitole - IRIT - 31062 Toulouse cedex 09 - Franck.Ravat@irit.fr

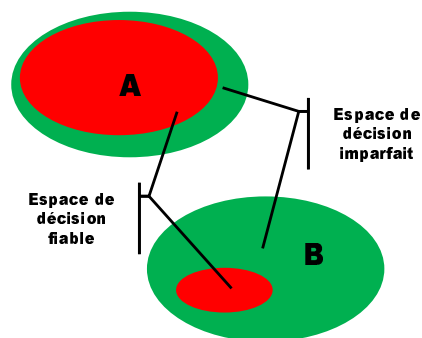
³ Université de Montpellier - UMR Espace-Dev - 34093 Montpellier Cedex 05 -
therese.libourel@umontpellier.fr

La vision standard des Systèmes d'Information d'Aide à la Décision (SIAD) repose sur d'une part, des sources transactionnelles classiques ou des répertoires de fichiers et, d'autre part, sur un ou plusieurs entrepôts de données à partir desquels sont effectuées des analyses décisionnelles reposant sur le principe OLAP (On-Line Analytical Processing), analyses restituant les indicateurs d'intérêt du domaine. Les entrepôts sont mis à jour régulièrement en extrayant les données des systèmes d'information transactionnels.

Un entrepôt de données étant matériellement indépendant du système d'information transactionnel, l'information y est structurée dans une vision d'analyse dans des cubes multidimensionnels dont les axes correspondent aux axes d'analyse des acteurs. Dans ces cubes multidimensionnels, l'information y est stockée avec comme objectif de réduire les temps de réponse afin qu'ils soient compatibles avec la prise de décision et de plus selon la granularité d'analyse souhaitée par les décideurs. Par exemple, sur la dimension spatiale, l'information peut être restituée à la parcelle, à la commune, au département, etc. Ces granularités sont en fait les échelles spatiales de décision pour de nombreux décideurs.

L'expérience montre que, dans beaucoup de domaines (pesticides, inondations, avalanches, etc.), les données environnementales stockées dans les systèmes transactionnels contiennent une part importante d'imperfection (Poujol et al., 2011; Vernier et al., 2013) et ce pour diverses raisons : changements de classe de précision des instruments de mesure de télédétection ou des capteurs, mise en place ou évolution des protocoles d'acquisition de données, amélioration de la précision de la géométrie des éléments du paysage, coût d'acquisition onéreux (satellites, enquêtes de terrain détaillées...), disponibilité hétérogènes de données cadastrales. De plus, les contraintes réglementaires comme la mise en place de la Directive européenne INSPIRE (<http://www.developpement-durable.gouv.fr/La-directive-europeenne-Inspire-de.html>) vont induire des problématiques d'imperfection au vu du nombre et de la disparité des secteurs concernés.

Dans le contexte environnemental, l'ensemble de l'information dont dispose le décideur relève de deux sous-ensembles que nous nommerons Espace de décision fiable et Espace de décision imparfait (cf. schéma). Lors d'une inondation, dans un contexte de gestion de crise, l'espace



intégrée au SIAD car elle est imparfaite.

décision fiable pourrait par exemple mobiliser le niveau du cours d'eau enrichi de la précision de la mesure mais cela n'est pas toujours possible. Dans le cas de données anciennes, de données obtenues via des techniques de crowdsourcing, etc., la mesure du niveau d'eau pourrait être disponible mais pas la précision. Cette information appartient à l'espace de décision imparfait. Bien que non conforme aux données de l'Espace de décision fiable, cette donnée a une valeur intrinsèque surtout si elle est rare (crue centenaire par exemple). Classiquement dans ce contexte, cette information incomplète ne serait pas

Tant que l'espace de décision fiable représente une part importante (cf. situation A), l'impact sur la décision finale reste acceptable. Il en est tout autrement lorsque l'espace de décision fiable est beaucoup plus petit que l'espace de décision imparfait (cf. situation B). En outre, l'expérience montre que le périmètre entre les deux sous-ensembles n'est pas constant et qu'il peut évoluer dans le temps dans un sens ou dans l'autre.

Classiquement, l'intégration de données sources dans les entrepôts de données vise à estimer les données manquantes et à éliminer ou à corriger les données imparfaites¹ afin d'obtenir un Système d'Information d'Aide à la Décision ne présentant a priori aucune imperfection. De ce fait, une grande quantité de données disponibles, dont certaines peuvent s'avérer stratégiques, est systématiquement exclue du processus décisionnel. Or dans de nombreux domaines, il peut s'avérer important, voire indispensable, d'intégrer des données imparfaites, parcellaires et d'actualité pour la prise de décision. De plus, la correction des données, lors du processus d'intégration des données décisionnelles, peut s'avérer fastidieuse, incomplète (spécifique à un nombre limité d'erreurs) et longue à mettre en œuvre.

L'approche que nous proposons est novatrice pour les Systèmes d'Information d'Aide à la Décision car nous ne souhaitons pas supprimer les données imparfaites mais les intégrer. Cette intégration au sein du processus d'entreposage des Systèmes d'Information d'Aide à la Décision soulève plusieurs questions de recherche :

- Comment enrichir la donnée par une sorte de « métadonnées » permettant d'explicitier la nature et la forme de l'imperfection ?
- Comment prendre en compte l'imperfection dès la conception des entrepôts de données mais probablement des systèmes d'information classiques ?
- Comment rendre plus flexible l'intégration des données imparfaites (donnée + métadonnée) ?
- Comment réaliser les agrégations aux différents niveaux de granularité sur des données imparfaites (donnée + métadonnée) ?
- Comment les imperfections de données influencent les outils OLAP l'exploration multidimensionnelle ?
- Comment rendre plus explicite la qualité des indicateurs présentés au décideur final ?

L'un des points clés reste la représentation et l'explicitation de la donnée imparfaite. Pour ce faire, il faut s'appropriier les connaissances des acteurs (qui elles aussi peuvent être imparfaites) mais aussi leur mode de raisonnement afin de mieux représenter ces données mais aussi la manière dont les acteurs les manipulent.

Cela peut conduire (i) à améliorer les techniques d'extraction d'informations à partir des données structurées, peu structurées ou pas structurées, (ii) à mobiliser des techniques d'évaluation et qualification des sources d'informations mais aussi des indicateurs calculés à partir de ces données, (iii) à prendre en compte des biais cognitifs dans les bases de connaissances imparfaites, etc.

1. **VERNIER, F., MIRALLES, A., PINET, F., CARLUER, N., GOUY, V., MOLLA, G. & PETIT, K.** (2013). EIS Pesticides: An environmental information system to characterize agricultural activities and calculate agro-environmental indicators at embedded watershed scales. *Agricultural Systems, Vol. 122*, pp. 11-21.
2. **POUJOL, G. & LABBE, S.** (2011, 07 juillet). *Fusion de données pour l'évaluation nationale du risque sismique : une approche préventive et générale des vulnérabilités*. SAGEO 2011 - International Conference on Spatial Analysis and GEOmatics, Paris, 8 + 1 poster.

¹ Il existe de nombreux travaux dont il serait difficile d'en faire état dans un résumé de deux pages.