

# Visualisation interactive de métadonnées pour aider les utilisateurs d'un logiciel de cartographie statistique à concevoir de meilleures cartes.

Perrine Pittet\*

\* Articque Software, 149 avenue Général de Gaulle, 37230 Fondettes  
ppittet@articque.com,  
<http://www.articque.com>

**Résumé.** CD7Online est l'application SaaS de la 7ème version de Cartes & Données (C & D), le logiciel de cartographie statistique décisionnelle et de géomarketing édité par Articque. C & D permet aux utilisateurs occasionnels de réaliser simplement des cartes à partir de données statistiques et géographiques. 25 ans de retours utilisateurs nous ont permis de voir que la qualité des cartes repose en partie sur la bonne connaissance des données dont disposent les utilisateurs et sur leur capacité à choisir des outils d'analyse et de représentation pertinents. Pour aider les utilisateurs à mieux comprendre leurs données et à réaliser des cartes de meilleure qualité, nous avons développé une brique sémantique avec un outil de visualisation interactif permettant de visualiser les connaissances extraites des espaces de travail des utilisateurs. Nous décrivons ici l'application CD7Online ainsi que l'outil de visualisation que nous présenterons lors de la démonstration logicielle.

## 1 Introduction

CD7Online est l'application SaaS de la 7ème version de Cartes & Données (2015) (C & D), le logiciel de cartographie statistique décisionnelle et de géomarketing édité par Articque. C & D permet aux utilisateurs de réaliser simplement des cartes à partir de données statistiques et géographiques. En effet, contrairement aux systèmes d'information géographiques et aux logiciels d'aide à la décision, les utilisateurs de CD7Online ne sont pas des spécialistes de la cartographie ou de la Business Intelligence (BI). C & D a été conçu dès le début dans le but d'être auto-explicatif, simple et intuitif pour des utilisateurs occasionnels. Néanmoins, certaines cartes créées par les utilisateurs sont parfois mal conçues ou ne reflètent pas le résultat escompté. Notre problématique est donc la suivante : comment aider les utilisateurs occasionnels d'un outil de cartographie statistique à construire des cartes qui répondent à leurs besoins dans un temps limité ? Les retours d'expérience des utilisateurs des différentes versions de C & D nous ont permis de voir que la pertinence des cartes repose, en partie, sur la bonne connaissance des données dont disposent les utilisateurs, et, sur leur capacité à choisir les bons outils d'analyse et de représentation. Pour répondre à la problématique, nous avons donc décidé de fournir aux utilisateurs les connaissances dont ils ont besoin pour comprendre rapidement et

exploiter au mieux leurs données. Nous avons développé un processus d’annotation sémantique automatisé de ces données afin d’extraire et de représenter ces connaissances. L’objectif principal est d’exploiter ces annotations pour permettre aux utilisateurs de visualiser, par l’intermédiaire d’un outil de visualisation graphique interactif, les concepts cartographiques liés à leurs données et les relations sémantiques qu’ils partagent (métadonnées). Cet outil leur permettra entre autres de choisir des données pertinentes pour construire des cartes pertinentes. C’est cet outil que nous avons choisi de décrire par le biais d’une démonstration du logiciel CD7Online. Dans ce papier nous décrivons dans un premier temps le logiciel et le processus de création de carte puis nous présentons le prototype de visualisation interactif.

## 2 CD7Online

### 2.1 Interface de création de carte

La création d’une carte dans CD7Online se fait au moyen d’un organigramme (cf. Fig. 1). Chaque icône représente un élément de la carte finale (une donnée statistique, un fond de carte, un traitement statistique, une représentation etc.). Les modules et leurs liaisons illustrent la méthode de construction et les étapes suivies pour obtenir le résultat final.

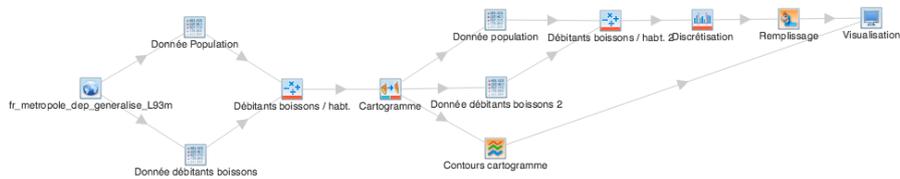


FIG. 1 – Organigramme d’un projet de carte dans CD7Online.

Dans l’interface de création de carte (cf. Fig. 2), la fenêtre de l’organigramme est la zone de création principale. Elle permet de modifier chaque paramètre de la carte. Les tableaux de données contiennent la liste des entités géographiques des territoires d’étude. Ils affichent les données importées ou calculées en rapport avec ces territoires. La fenêtre de visualisation affiche le résultat cartographique de l’organigramme et permet de faire la mise en page du document pour l’export de la carte. Différents types de modules sont utilisables dans l’organigramme depuis le bandeau. Les modules sources permettent d’importer les fonds de carte et les données statistiques, qui sont stockées par l’utilisateur dans ses espaces de travail. Les modules opérateurs comprennent : les opérateurs géographiques dont les traitements concernent les fonds de carte et l’aspect spatial du projet, les opérateurs statistiques qui opèrent les traitements statistiques sur les données, et les opérateurs mixtes dont les traitements font les deux. Les modules représentation correspondent aux méthodes graphiques pour représenter les informations sur la carte. Elles sont adaptées à l’illustration des différents types de phénomènes et doivent être choisies en fonction.

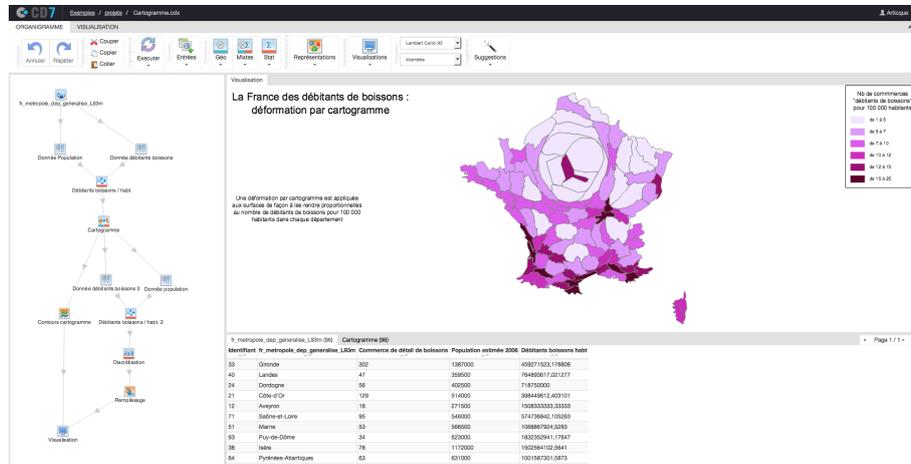


FIG. 2 – Interface de création de carte dans CD7Online.

## 2.2 Processus de création de carte

La création d'un organigramme passe par cinq grandes étapes directrices qui suivent la logique de réalisation d'une analyse cartographique. (1) Choix du fond de carte : cette étape permet de définir l'espace et le niveau géographique (niveau de découpage territorial) à étudier et sur lesquels les informations vont être représentées. (2) Choix des données statistiques : cette étape nécessite de connaître au préalable l'espace et le niveau géographique du fond de carte afin de sélectionner des données compatibles avec le fond de carte. Par exemple si le fond de carte est celui de la France par commune, les données doivent être calculées et identifiées par un code commune français. Aussi les données statistiques doivent être compatibles entre elles, c'est-à-dire porter sur des populations, des indicateurs et des périodes comparables. (3) Modification/transformation des sources : c'est à cette étape que l'utilisateur choisit les opérateurs de transformation (statistiques, géographiques et/ou mixtes) pour filtrer, analyser et traiter les données pour les rendre représentables. Il va pouvoir par exemple choisir un opérateur statistique pour calculer une répartition sur plusieurs séries de données. Le choix des opérateurs est conditionné par le type (données continues, qualitatives ou discrètes) et le nombre (une à plusieurs) des données sources. (4) Choix des représentations : à cette étape le choix des représentations est conditionné par les opérateurs de transformation utilisés en amont. En effet, en fonction du type des résultats en sortie d'un opérateur de transformation et du phénomène observé, certaines représentations conviendront et d'autres non. Par exemple, le résultat d'une discrétisation va nécessiter une représentation à base de remplissage et de dégradés de couleurs. (5) Visualisation et organisation de la carte : cette étape permet de visualiser le résultat cartographique des différentes étapes en amont et d'éditer certains aspects visuels avant d'exporter la carte.

Nous voyons ici que, bien que le processus de création de carte soit simple et bien défini pour guider l'utilisateur dans C & D, la plupart des étapes nécessite tout de même une bonne connaissance du contenu des données sources et quelques notions de base d'analyse statistique

## Visualisation de métadonnées pour guider la création de cartes statistiques

et de représentation cartographique. Afin d'apporter ces connaissances à l'utilisateur au moment de la réalisation d'une carte, nous avons conçu et intégré une brique sémantique dans CD7Online.

### 2.3 La Brique Sémantique

La brique sémantique de CD7Online comporte deux composants majeurs : une ontologie d'application et un annotateur sémantique. L'ontologie CD7, dont le développement et les concepts principaux sont décrits respectivement dans Pittet et Barthélémy (2015a) et Pittet et Barthélémy (2015b), modélise les connaissances du domaine d'application de CD7Online dont : les fichiers de données statistiques (tableaux de données), les fichiers de données géographiques (fonds de cartes) et les projets de cartes (organigrammes), mais aussi les processus métiers liés à l'analyse statistique et à la cartographie. L'ontologie est spécifiée dans le langage OWL DL McGuinness et al. (2004) et a été développée pour supporter l'annotation sémantique automatisée des données utilisateurs (sources et projets de cartes) et un système de recommandation. La tâche d'annotation sémantique automatique des données utilisateurs, dont le processus étend la méthodologie de Hignette (2007), permet l'extraction des métadonnées (niveaux géographiques, les indicateurs statistiques, etc.) à partir des données de l'utilisateur depuis ses espaces de travail pour représenter leurs relations avec ces données via des annotations RDF Klyne et Carroll (2006). Durant cette tâche, les contenus des données sont lemmatisés et des mesures de similarité lexicales s'appuyant sur les termes et les poids des concepts spécifiés dans l'ontologie sont effectuées afin d'identifier les métadonnées. Dans la démonstration nous nous focaliserons sur l'outil permettant de visualiser les métadonnées des données utilisateur décrit ci-après.

## 3 L'outil de visualisation des métadonnées

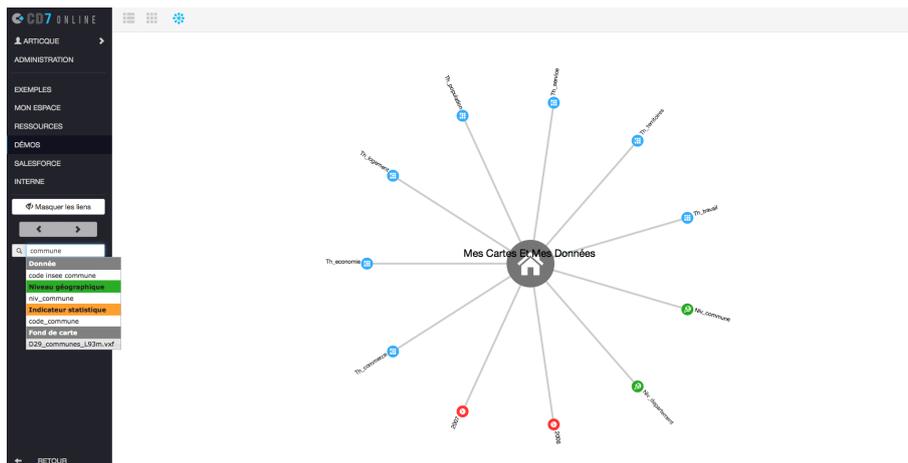


FIG. 3 – Interface de l'outil de visualisation interactif.

Construit au dessus du triplestore stockant les annotations sémantiques des données utilisateur, un outil de visualisation, présenté sous la forme d'un graphe radial de type Reingold Tilford interactif, permet aux utilisateurs de parcourir intuitivement leurs espaces de travail au travers des annotations sémantiques. L'avantage de ce type de représentation est qu'il rend possible la visualisation d'un grand nombre de données dans un espace restreint et sans occlusion Herman et al. (2000). Intégré à la manière des autres navigateurs de documents dans le logiciel, cet outil permet deux types de navigation : une navigation à la main dans le graphe lui-même (cf. Fig. 3 partie droite) et une navigation via un champ de recherche autocomplété (cf. Fig. 3 partie gauche). Ces deux navigations peuvent être utilisées conjointement.

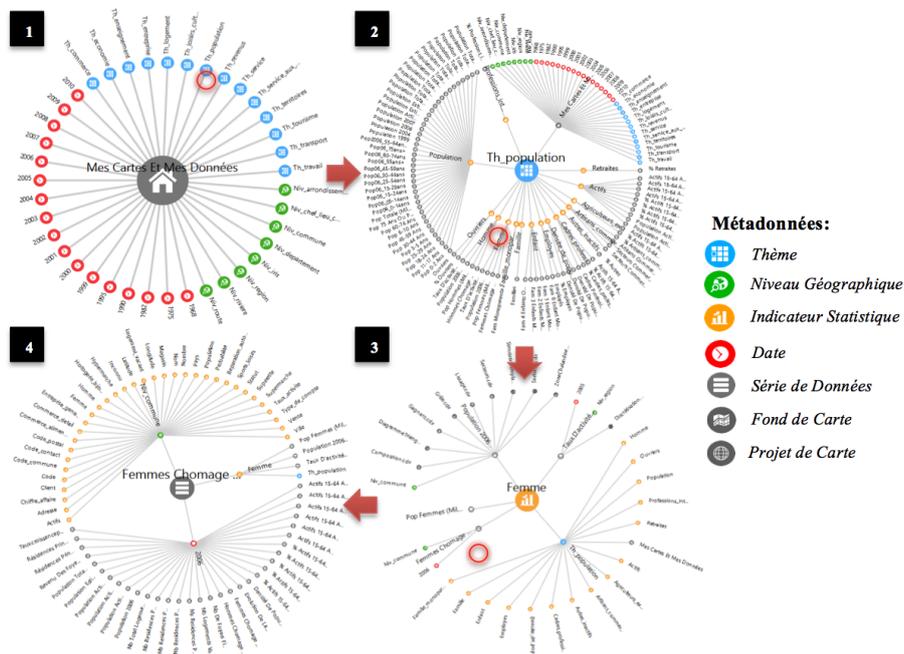


FIG. 4 – Exemple de navigation dans l'outil de visualisation de métadonnées.

La figure 4 décrit un exemple de processus de navigation à la main dans le graphe. La première étape montre la visualisation par défaut du graphe sur lequel l'utilisateur peut voir les principales métadonnées identifiées (thèmes, dates et niveaux géographiques) liées aux données de son espace de travail (cf. Fig. 4 étape 1). Si l'utilisateur recherche une série de données liées à la population, il peut alors cliquer sur le thème « th\_population ». Le graphique se transforme alors pour afficher les indicateurs statistiques relatifs à ce thème (cf. Fig. 4 étape 2). Parmi ceux-ci, l'utilisateur clique sur l'indicateur « Femme », parce que la série de données qu'il recherche est liée à une population de femmes. Le graphique se transforme encore pour afficher les séries de données statistiques correspondant à cet indicateur (cf. Fig. 4 étape 3). L'utilisateur trouve la série de données qu'il cherchait, « Femmes chômage ». En cliquant dessus, il peut voir que cette série de données couvre les statistiques sur les femmes au chômage

en France au niveau géographique « fr commune » (ville) sur l'année « 2006 » (cf. Fig. 4 étape 4). Il peut la sélectionner et ensuite continuer à naviguer dans le graphe pour trouver d'autres séries de données qui sont compatibles. Par exemple d'autres données sur le même niveau géographique, d'autres avec le même indicateur statistique mais sur une année différente, etc. On aurait pu imaginer la même navigation à partir du champ de recherche autocomplété. L'utilisateur aurait alors commencé à taper le terme « chômage », pour obtenir la liste des séries de données contenant le terme « chômage », puis aurait cliqué sur la série « Femme Chômage » pour accéder à la vue correspondant à l'étape 4. Enfin, les séries de données peuvent être importées dans l'organigramme d'un nouveau projet de carte que l'utilisateur peut alors compléter pour réaliser sa carte.

## Références

- Herman, I., G. Melançon, et M. S. Marshall (2000). Graph visualization and navigation in information visualization : A survey. *Visualization and Computer Graphics, IEEE Transactions on* 6(1), 24–43.
- Hignette, G. (2007). *Annotation sémantique floue de tableaux guidée par une ontologie*. Ph. D. thesis, AgroParisTech.
- Klyne, G. et J. J. Carroll (2006). Resource description framework (rdf) : Concepts and abstract syntax.
- McGuinness, D. L., F. Van Harmelen, et al. (2004). Owl web ontology language overview. *W3C recommendation* 10(10), 2004.
- Pittet, P. et J. Barthélémy (2015a). An application ontology to help users of a geo-decision software understanding their data.
- Pittet, P. et J. Barthélémy (2015b). Experience of formal application ontology development to enhance user understanding in a geo business intelligence saas platform. In *Formal Ontologies Meet Industry*, pp. 51–62. Springer.

## Summary

CD7Online is the SaaS application of the seventh version of Cartes & Données (C & D), a decisional statistical mapping and geomarketing software edited by Articque. It allows casual users to easily create maps from statistical and geographical data. 25 years of C & D users' feedbacks allowed us to see that the quality of maps is partly based on the users' good knowledge regarding their data and their ability to select relevant analysis and representation tools. To help users better understand their data and realize better maps, we developed a semantic layer on the top of CD7Online, with an interactive visualization tool for browsing user data knowledge extracted from their workspaces. Here we describe the CD7Online application and the visualization prototype that we will present at the demo session.