



Faciliter la publication dans le Web des données

26 mars, 2013



<http://datalift.org>



Datalift en bref

- **Une plate-forme de production et de publication de données liées**
- **Projet open-source, téléchargeable et opérationnelle**
- **Cofinancé par l'ANR pendant 3 ans et demi (09/2010 – 03/2014)**



« Pub-lier » des données avant Datalift

- Beaucoup d'expertise et des outils à rassembler
 - Une pile technologique un peu compliquée à maîtriser
 - Une connaissance des vocabulaires et référentiels partagés
- Une démarche trop souvent isolée
 - On publie *ses propres données avec ses propres vocabulaires*
 - Les données sont ouvertes, le format est partagé, mais elles restent souvent dans un **silosémantique**




Partenariat solide et crédible

- Enseignement et recherche
 - EURECOM (Institut Mines-Telecom)
 - INRIA (EXMO, WIMMICS)
 - LIRMM (Université de Montpellier)
- Industriels
 - Atos (intégrateur)
 - Mondeca (fournisseur de solutions sémantiques depuis 2000)
- Fournisseurs de données
 - IGN (données géographiques)
 - Insee (données statistiques, nomenclatures)
- Secteur associatif et collectivités
 - FING (Fondation Internet Nouvelle Génération)



Des modules pour chaque « étage de l'ascenseur »

- Publication via SPARQL end-point (public ou non)
 - Visualisation, navigation, requêtage, alignements ...
 - Transformations diverses sur les données RDF
 - Réconciliation et interconnexion des entités partagées (LOD)
 - Réconciliation avec des vocabulaires partagés (LOV)
 - Conversion des formats sources (CSV, SQL, XML) en RDF
- 

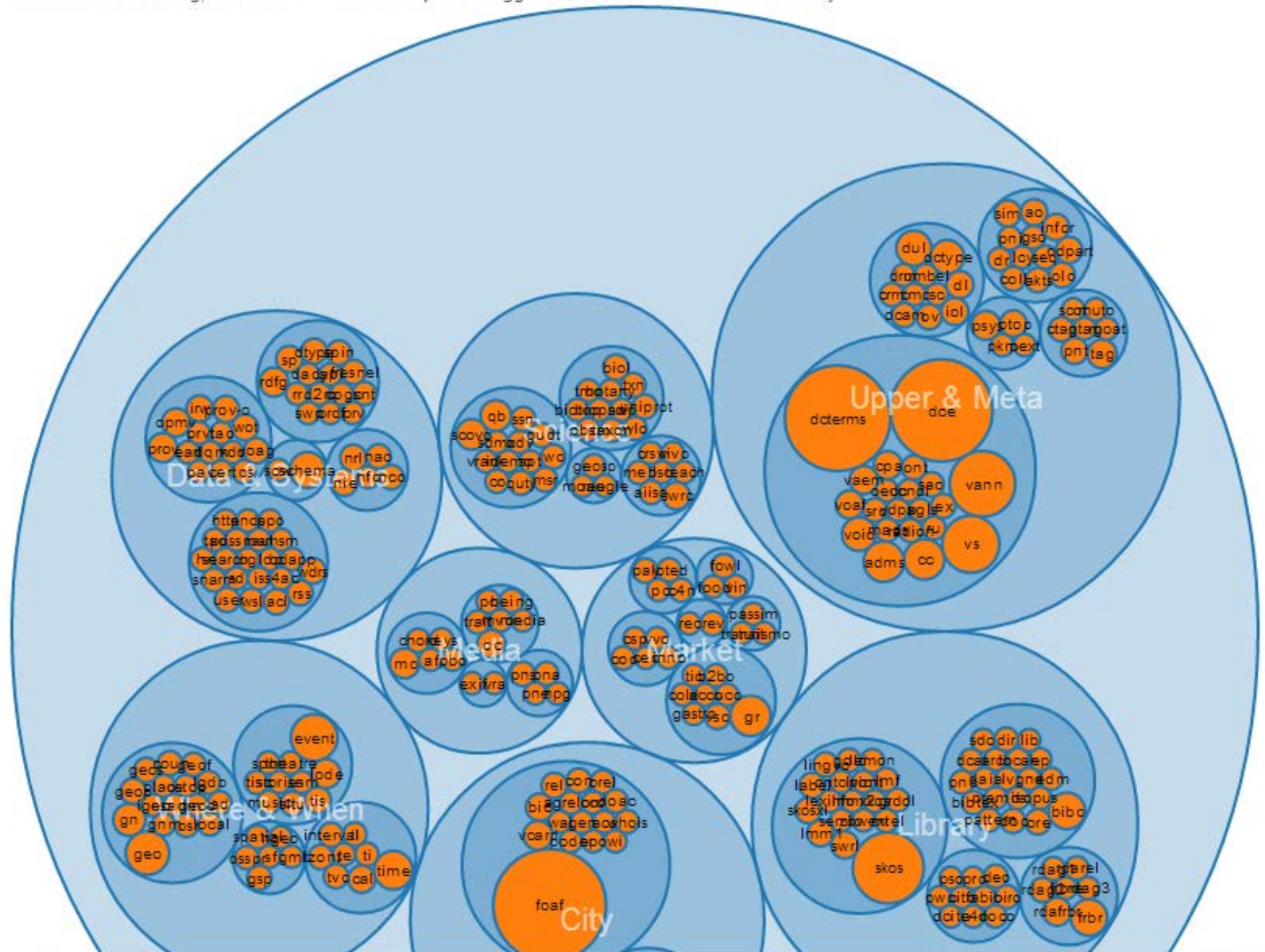


Vocabulaires qualifiés et partagés

- **Linked Open Vocabulaires**
 - Catalogue de vocabulaires réutilisables pour la description des données
 - Ontologies en RDFS/OWL
 - Aujourd'hui hébergé par Open Knowledge Foundation
- **Description innovante des vocabulaires**
 - Détection et représentation des liens et dépendances
 - Recherche fédérée, stockage des versions
 - Monitoring semi-automatique de l'écosystème des vocabulaires
- **Contenu intégré dans la plate-forme Datalift**
 - Recherche de vocabulaires, alignement ...

Vocabulary spaces (325):

Classifying vocabularies is a brand new challenge for Library Science, and no existing classification scheme seems to fit the need so far. Unless we missed something, and in that case we are open to suggestions from the librarian community.





Fournisseurs de données de référence

- Insee et IGN
 - Données géographiques et statistiques
 - Codes et nomenclatures
- Cas d'école pour l'interopérabilité
 - Objets communs : entités géographiques
 - Points de vue différents sur ces objets
territoire administratif, statistique, géométrique...
- Un travail coopératif entre ces deux acteurs
 - Politique commune sur les URI
 - Rapprochement des entités et des vocabulaires



Données de références déjà publiées

- Par l'IGN
 - Ontologie topographique et données administratives
 - A suivre sur data.ign.fr
- Par l'Insee
 - Code Officiel Géographique
 - Codes et Nomenclatures (NAF)
 - Données de population (recensement)
 - A suivre sur data.insee.fr

Un exemple : collèges de la Gironde

UAI	ADR_L1	ADR_L4VOIE	ADR_L6CP	ADR_L6COMMUNE	INSEE_COMMUNE
0332723F	COLLEGE EMILE DURKHEIM	6 RUE LA MICHERE	33240	PEUJARD	33321
0332743C	COLLEGE ALIENOR D AQUITAINE	AVENUE DES SAPINETTES	33127	MARTIGNAS SUR JALLE	33273
0332746F	COLLEGE EMILE COMBES	76 RUE EMILE COMBES	33000	BORDEAUX	33063
0332768E	COLLEGE ALIENOR D AQUITAINE	16 RUE DON DEVIENNE	33800	BORDEAUX	33063
0332806W	COLLEGE LE MIRAIL	61 COURS DE LA SOMME	33800	BORDEAUX	33063
0332934K	COLLEGE LE PIAN-SUR-GARONNE	1 RUE LALIEYRE NORD	33490	LE PIAN SUR GARONNE	33323
0332982M	COLLEGE LE TEICH	AVENUE F. MITTERAND	33470	LE TEICH	33527
0333093H	COLLEGE LEONARD DE VINCI	AVENUE DU LYCEE	33160	ST AUBIN DE MEDOC	33376
0333108Z	COLLEGE JEAN MOULIN	RUE ALI CHEKKAL	33492	LE BOUSCAT CEDEX	33069
0333121N	COLLEGE GASTON FLAMENT	33 AVENUE LEON DELAG...	33380	MARCHEPRIME	33555
0333129X	COLLEGE SAINT JOSEPH	12 AVENUE DU 8 MAI	33291	BLANQUEFORT CEDEX	33056
0333132A	COLLEGE CARBON-BLANC	RUE DES FUTAIES PLAINE ...	33564	CARBON BLANC CEDEX	33096
0331532L	COLLEGE JEANNE D ARC	3 PLACE DE LA FRATERNI...	33230	ST MEDARD DE GUIZIERES	33447
0331617D	COLLEGE CLAUDE MASSE	4 AVENUE PIERRE BARRE	33440	AMBARES ET LAGRAVE	33003
0331618E	COLLEGE LEONARD LENOIR	16 RUE L LENOIR	33015	BORDEAUX CEDEX	33063
0331619F	COLLEGE GEORGES LAPIERRE	2 RUE PIERRE BROSSELE...	33305	LORMONT CEDEX	33249
0331620G	COLLEGE CAMILLE CLAUDEL	RUE DU STADE	33360	LATRESNE	33234
0331621H	COLLEGE HENRI DE NAVARRE	RUE JEAN ZAY	33230	COUSTRAS	33138
0331622J	COLLEGE FONTAINES DE MONJOURS	7 RUE F DE MONJOURS	33173	GRADIGNAN CEDEX	33192
0331623K	COLLEGE ALOUETTE		33601	PESSAC CEDEX	33318
0331644H	COLLEGE ST JEAN	7 RUE MAQUIS VIGNES O...	33250	PAUILLAC	33314
0331662C	COLLEGE ALAIN FOURNIER	83 RUE SAINT GENES	33023	BORDEAUX CEDEX	33063
0331663D	COLLEGE CHEVERUS	62 RUE FIEFFE	33800	BORDEAUX	33063
0331664E	COLLEGE FRANCOIS MAURIAC	131 AVENUE MONTAIGNE	33166	ST MEDARD EN JALLES C...	33449
0331666G	COLLEGE ALIENOR D AQUITAINE	22 ROUTE DU MARTINET	33770	SALLES	33498
0331419N	COLLEGE GEORGES RAYET	42 CHEMIN CORNIER	33271	FLOIRAC CEDEX	33167
0331433D	COLLEGE PIERRE MARTIN	RUE DU COLLEGE	33420	RAUZAN	33350
0331435F	COLLEGE LES EYQUEMS	RUE TOUR DE VEYRINES	33700	MERIGNAC	33281
0331461J	COLLEGE CASSIGNOL	62 RUE LAROCHE	33000	BORDEAUX	33063
0331462K	COLLEGE FRANCISCO GOYA	56 RUE DU CDT ARNOULD	33023	BORDEAUX CEDEX	33063
0331464M	COLLEGE JEAN ZAY	AVENUE JEAN ZAY	33151	CENON CEDEX	33119
0331490R	COLLEGE LESTONNAC	92 ROUTE DE FARGUES	33360	CARIGNAN DE BORDEAUX	33099



Des données « mûres » pour le lifting

- Format non propriétaire (CSV)
- Un type d'objets bien défini (Collège)
- Des propriétés simples
 - identifiant, nom, adresse, commune ...
- Des identifiants partagés
 - code établissement, code commune
- La transformation de ces données en RDF peut être en grande partie automatisée

3 cellules = un triplet RDF

UAI	ADR_L1	ADR_L4VOIE	ADR_L6CP	ADR_L6COMMUNE	INSEE_COMMUNE
0332723F	COLLEGE EMILE DURKHEIM	6 RUE LA MICHERE	33240	PEUJARD	33321
0332743C	COLLEGE ALIENOR D AQUITAIN	AVENUE DES SAPINETTES	33127	MARTIGNAS SUR JALLE	33273
0332746F	COLLEGE EMILE COMBES	76 RUE EMILE COMBES	33000	BORDEAUX	33063
0332768E	COLLEGE ALIENOR D AQUITAIN	16 RUE DON DEVIENNE	33800	BORDEAUX	33063
0332806W	COLLEGE LE MIRAIL	61 COURS DE LA SOMME	33800	BORDEAUX	33063
0332934K	COLLEGE LE PIAN-SUR-GARONNE	1 RUE LALIEYRE NORD	33490	LE PIAN SUR GARONNE	33323
0332982M	COLLEGE LE TEICH	AVENUE F. MITTERAND	33470	LE TEICH	33527
0333000H	COLLEGE LEONARD DE VINCI	AVENUE DU LYCEE	33160	ST AUBIN DE MEDOC	33070
0333108Z	COLLEGE JEAN MOULIN	RUE ALI CHEKKAL	33492	LE BOUSCAT CEDEX	33069
0333121N	COLLEGE GASTON FLAMENT	33 AVENUE LEON DELAG...	33380	MARCHEPRIME	33555
0333129X	COLLEGE SAINT JOSEPH	12 AVENUE DU 8 MAI	33291	BLANQUEFORT CEDEX	33056
0333132A	COLLEGE CARBON-BLANC	RUE DES FUTAIES PLAINE ...	33564	CARBON BLANC CEDEX	33096
0331532L	COLLEGE JEANNE D ARC	3 PLACE DE LA FRATERNI...	33230	ST MEDARD DE GUIZIERES	33447
0331617D	COLLEGE CLAUDE MASSE	4 AVENUE PIERRE BARRE	33440	AMBARES ET LAGRAVE	33003
0331618E	COLLEGE LEONARD LENOIR	16 RUE L LENOIR	33015	BORDEAUX CEDEX	33063
0331619F	COLLEGE GEORGES LAPIERRE	2 RUE PIERRE BROSSOLE...	33305	LORMONT CEDEX	33249
0331620G	COLLEGE CAMILLE CLAUDEL	RUE DU STADE	33360	LATRESNE	33234
0331621H	COLLEGE HENRI DE NAVARRE	RUE JEAN ZAY	33230	COUSTRAS	33138
0331622J	COLLEGE FONTAINES DE MONJOUS	7 RUE F DE MONJOUS	33173	GRADIGNAN CEDEX	33192
0331623K	COLLEGE ALOUETTE		33601	PESSAC CEDEX	33318
0331644H	COLLEGE ST JEAN	7 RUE MAQUIS VIGNES O...	33250	PAUILLAC	33314
0331662C	COLLEGE ALAIN FOURNIER	83 RUE SAINT GENES	33023	BORDEAUX CEDEX	33063
0331663D	COLLEGE CHEVERUS	62 RUE FIEFFE	33800	BORDEAUX	33063
0331664E	COLLEGE FRANCOIS MAURIAC	131 AVENUE MONTAIGNE	33166	ST MEDARD EN JALLES C...	33449
0331666G	COLLEGE ALIENOR D AQUITAIN	22 ROUTE DU MARTINET	33770	SALLES	33498
0331419N	COLLEGE GEORGES RAYET	42 CHEMIN CORNIER	33271	FLOIRAC CEDEX	33167
0331433D	COLLEGE PIERRE MARTIN	RUE DU COLLEGE	33420	RAUZAN	33350
0331435F	COLLEGE LES EYQUEMS	RUE TOUR DE VEYRINES	33700	MERIGNAC	33281
0331461J	COLLEGE CASSIGNOL	62 RUE LAROCHE	33000	BORDEAUX	33063
0331462K	COLLEGE FRANCISCO GOYA	56 RUE DU CDT ARNOULD	33023	BORDEAUX CEDEX	33063
0331464M	COLLEGE JEAN ZAY	AVENUE JEAN ZAY	33151	CENON CEDEX	33119
0331490R	COLLEGE LESTONNAC	92 ROUTE DE FARGUES	33360	CARIGNAN DE BORDEAUX	33099

... mais la syntaxe n'est pas tout!

- Le « triplet » extrait automatiquement ...

sujet

0333108Z

prédicat

INSEE_COMMUNE

objet

'33069'

- Est-il utilisable en l'état? Pas vraiment ...

- Ce n'est pas encore tout à fait du RDF (il manque les URI)
- Il y a beaucoup d'implicite ... à expliciter

- Ce qui reste à expliciter

- Toutes les lignes décrivent un même type d'objets : des collègues
- Identifiés par des codes uniques définis sans doute ailleurs
- La signification de chaque prédicat dans un vocabulaire partagé



Expliciter entités et vocabulaires

sujet

0333108Z

prédictat

INSEE_COMMUNE

objet

'33069'

- Traduire le sujet en URI de publication
<http://data.datalocale.fr/equipement/college/0333108Z>
- Traduire le prédicat en URI de référence
http://www.w3.org/2003/01/geo/wgs84_pos#location
- Traduire l'objet en URI de référence
<http://id.insee.fr/geo/commune/33069>
- Expliciter le type du sujet dans un vocabulaire
<http://data.datalocale.fr/org#CollegePublic>

Le projet dans Datalift

Projets

[Geonames](#)

[Datalocale -
Collèges](#)

[Environnement](#)



[Nouveau
projet](#)

Description

[Sources](#)

[Ontologies](#)

Description : Données collèges du CG 33
Licence : [Attribution, 3.0 non transposé \(CC BY 3.0\)](#)
Dernière modification le 22 mars 2013 15:49:13
Créé le 4 oct. 2012 13:16:04 par Bernard

Modules de transformation applicables

1. [Transformation directe CSV vers RDF](#)
2. [Chargement d'une source RDF fichier](#)
3. [RDF to RDF transformation \(CONSTRUCT\)](#)
4. [Renommage des URI RDF](#)
5. [Publication vers le RDF store public](#)
6. [Export des données vers fichier RDF](#)
7. [Export des données vers fichier CSV](#)
8. [Interconnexion avec Silk](#)
9. [Transformation des Strings en URIs](#)

Notre collège « lifté »

Sujet ↕	Prédicat	Objet
http://data.datalocale.fr/equipement/College-0333108Z	http://www.w3.org/2000/01/rdf-schema#label	"COLLEGE JEAN MOULIN"@fr
http://data.datalocale.fr/equipement/College-0333108Z	http://data.datalocale.fr/org#nomLong	"LE BOUSCAT jean moulin"@fr
http://data.datalocale.fr/equipement/College-0333108Z	http://data.datalocale.fr/org#nomCourt	"jean moulin"@fr
http://data.datalocale.fr/equipement/College-0333108Z	http://schemas.talis.com/2005/address/schema#streetAddress	"RUE ALI CHEKKAL"
http://data.datalocale.fr/equipement/College-0333108Z	http://schemas.talis.com/2005/address/schema#postalCode	"33492"
http://data.datalocale.fr/equipement/College-0333108Z	http://schemas.talis.com/2005/address/schema#localityName	"LE BOUSCAT CEDEX"
http://data.datalocale.fr/equipement/College-0333108Z	http://www.w3.org/2003/01/geo/wgs84_pos#location	http://rdf.insee.fr/geo/COM_33069
http://data.datalocale.fr/equipement/College-0333108Z	http://www.w3.org/1999/02/22-rdf-syntax-ns#type	http://data.datalocale.fr/org#CollegePublic
http://data.datalocale.fr/equipement/College-0333108Z	http://rdf.insee.fr/def/territoire/bpe#secteur	http://id.insee.fr/codes/territoire/Secteur/Public
http://data.datalocale.fr/equipement/College-0333108Z	http://data.datalocale.fr/org#telephone	"0556089026"
http://data.datalocale.fr/equipement/College-0333108Z	http://data.datalocale.fr/org#fax	"0556029369"
http://data.datalocale.fr/equipement/College-0333108Z	http://data.datalocale.fr/org#mail	"ce.0333108Z@ac-bordeaux.fr"
http://data.datalocale.fr/equipement/College-0333108Z	http://data.datalocale.fr/org#numZap	"02"
http://data.datalocale.fr/equipement/College-0333108Z	http://data.datalocale.fr/org#numSecteur	"0330215"
http://data.datalocale.fr/equipement/College-0333108Z	http://rdf.insee.fr/def/territoire/bpe#caracterePresent	http://data.datalocale.fr/org#CaractereDP
http://data.datalocale.fr/equipement/College-0333108Z	http://data.datalocale.fr/org#dateOuverture	"2004-09-01"
http://data.datalocale.fr/equipement/College-0333108Z	http://www.w3.org/2003/01/geo/wgs84_pos#lat	"44.863354"
http://data.datalocale.fr/equipement/College-0333108Z	http://www.w3.org/2003/01/geo/wgs84_pos#long	"-0.60558516"
http://data.datalocale.fr/equipement/College-0333108Z	http://rdf.insee.fr/def/territoire/bpe#anneeDescription	"2011"
http://data.datalocale.fr/equipement/College-0333108Z	http://www.w3.org/1999/02/22-rdf-syntax-ns#type	http://rdf.insee.fr/def/territoire/bpe#EquipementScol

Visualisation

Explorer

Hiérarchie

Facettes

Territoire administratif > Canton

- Arrondissement
- Arrondissement municipal
- Canton**
- Collectivité territoriale
- Commune
- Département
- EPCI
- Région

Canton

Arcachon

Arcachon

<http://id.insee.fr/geo/canton/3301>

CANTON

Autres propriétés

chef-lieu	<input checked="" type="radio"/> Arcachon
subdivision directe	<input checked="" type="radio"/> Arcachon
code canton	3301
code INSEE	3301
subdivision de	http://id.insee.fr/geo/arrondissement/336 - Arcachon
sameAs	http://data.ign.fr/id/geofla/canton/3301
type	Canton
latitude	44.65
longitude	-1.16667

<input checked="" type="radio"/> COLLEGE MARIE BARTETTE	location
<input checked="" type="radio"/> COLLEGE ST ELME	
<input checked="" type="radio"/> CITY STADE	
<input checked="" type="radio"/> PLATEAU SPORTIF DU COLLEGE M.BARTETTE	
<input checked="" type="radio"/> STADE JEAN BROUSSE	
<input checked="" type="radio"/> PISTE DE SKATE	
<input checked="" type="radio"/> GOLF INTERNATIONAL	



Merci de votre attention !



Alexander Polonsky
Directeur Marketing

3 Cité Nollez
75018 Paris, France
+33 1 44 92 35 03

alexander.polonsky@mondeca.com
www.mondeca.com