

Projet d'article :

« l'ontologie officielle française du tourisme en France - v2.0 :

une avancée structurelle fondamentale et fondatrice d'un renouveau possible en matière d'observation, de mesure et d'analyse des données touristiques, annonciatrice de nouvelles modalités de gestion pour un développement futur d'un tourisme plus résilient et durable » au lendemain de la crise du Covid-19.

Résumé :

Sous l'égide de la DGE, l'ontologie nationale du tourisme a été développée par DATAtourisme dans le cadre d'un projet collaboratif impliquant potentiellement l'ensemble des parties prenantes du domaine touristique. Elle repose sur le concept central de « **points d'intérêts touristiques** " (*POI = Point of interest*) ; ce concept, à large spectre, regroupe notamment les divertissements et événements, sites naturels et culturels, loisirs et activités sportives, produits touristiques, visites, hébergements, boutiques, restaurants, etc. Les données touristiques sont regroupées dans un format de partage commun, organisé sous forme de réseau, structuré sur plusieurs niveaux et l'ontologie est complétée par un thésaurus.

Les données ainsi classifiées et décrites sont extraites de différentes bases de données officielles françaises recensés par les Offices de Tourisme, Agences Départementales et Comités Régionaux du Tourisme, fêtes et manifestations... ainsi que par les gestionnaires de sites patrimoniaux – naturels – culturels, d'activités de loisirs et d'animation, d'itinéraires, d'hébergements, de boutiques, de restaurants etc.

Ce format de référence est notamment utilisé par la plateforme nationale DATAtourisme pour la diffusion en OpenData des données touristiques. En sa qualité de projet collaboratif, DATAtourisme est destiné à s'enrichir des contributions de ses usagers (<http://www.datatourisme.fr> et <https://framagit.org/datatourisme/ontology/tree/master>).

Rappels génériques importants :

a) Définition d'une ontologie : « dans son sens le plus général, le concept d'ontologie interroge sur la signification du mot « être » : qu'est-ce que l'être, qu'est-ce qui existe ? ». C'est une question considérée comme « inaugurale », c'est-à-dire première dans le temps et première dans l'ordre de la connaissance. Une fois clarifié, le concept d'ontologie peut se décliner en nomenclatures et classifications, à visées plus pratiques et opérationnelles sur le terrain.

Considérée à juste titre comme très difficile, la définition d'une ontologie, notamment pour le domaine du tourisme, constitue « la tâche princeps », indispensable pour que toutes les parties prenantes puissent savoir de quoi l'on parle dans le monde du tourisme, connaître ce qui existe en matière de sites, de ressources et d'activités touristiques et disposer d'un vocabulaire commun pour gérer – échanger – développer ce patrimoine et ces activités.

L'ontologie développée par DATAtourisme comporte 500 entités, dont 372 classes de ressources et d'activités, 61 propriétés d'objets, 50 propriétés des données, 8 propriétés des annotations et 9 types des données.

Pour l'ensemble des activités touristiques, l'ontologie développée par DATAtourisme se réfère à la classification APE « **NAF Rév. 2** » de l'INSEE (*dernière mise à jour le 04/03/2019*) : <https://www.insee.fr/fr/information/2406147>. La **NAF, nomenclature d'activités française**, est une nomenclature des activités économiques productives, principalement élaborée pour faciliter l'organisation de l'information économique et sociale. Afin de faciliter les comparaisons internationales, elle a la même structure que la **nomenclature d'activités européenne NACE**, elle-même dérivée de la **nomenclature internationale CITI**.

Pour toutes les instances territoriales, le **COG (code officiel géographique)** de l'INSEE sert de référence (*dernière mise à jour le 25/02/2020*) : <https://www.insee.fr/fr/information/2560452>

Plus en détail pour les activités d'hébergement collectif commercial, l'ontologie développée par DATAtourisme se fonde sur l'ontologie autrichienne ACCO développée en 2013 et accessible par le lien suivant : <http://ontologies.sti-innsbruck.at/acco/ns.html>

b) Définition d'un thésaurus documentaire : Un thésaurus, thésaurus de descripteurs ou thésaurus documentaire, est une liste organisée de termes contrôlés et normalisés représentant les concepts d'un domaine de la connaissance.

Au 20 avril 2020, le thésaurus de DATAtourisme comporte 1.107 instances, 415 labels de classement des POI et 51 classes parentes.

<http://www.datatourisme.fr/resource/core/2.0#>.

Appliquée au domaine touristique, cette définition du thésaurus entraîne plusieurs conséquences importantes :

- 1) **La liste des termes utilisés est organisée** : un thésaurus est structuré comme un réseau de nœuds reliés par des relations ; ces nœuds et relations doivent être reconnus et acceptés par l'ensemble des parties prenantes ou intéressées et répondre aux réalités opérationnelles sur le terrain ;
- 2) **Les termes sont contrôlés** : il existe donc une instance de contrôle et de validation pour ces termes, en l'occurrence, DATAtourisme ;
- 3) **Les termes sont normalisés** : ils s'inscrivent dans des catégories qui doivent présenter les principales caractéristiques suivantes :
 - a. Couvrir tout le champ des ressources et des offres touristiques (concept de POI et de leurs sous classes) ; (*pas de lacunes*) ;
 - b. Ne pas se chevaucher ; (*pas de redondance*) ;
 - c. Respecter le principe général d'intégrité ;
- 4) La liste des termes ainsi que la structure du réseau présidant à leur organisation et mise en relation pourra s'adapter par consensus en fonction des évolutions des ressources et des offres ; elle présente donc un intéressant caractère « **dynamique et adaptatif** », fondamental pour la pérennité du système. Ils sont issus de quatorze ontologies externes.

c) Définition d'un graphe : dans le domaine de la théorie des graphes, un graphe est une structure composée d'objets dont certaines paires sont en relation. Les graphes sont des modèles de réseaux reliant des objets au moyen de relations.

Dans le domaine du tourisme, DATAtourisme a développé une base de données orientée « objet » (POI) utilisant la théorie des graphes, donc avec des nœuds et des arcs de mise en relation, permettant de représenter et de stocker les données relatives aux ressources et aux offres touristiques, tout en explicitant leurs liens et relations de manière normalisée et permanente.

Il s'agit donc d'un système de gestion de bases de données relationnelle, orientée « utilisateurs », c'est-à-dire destinée à pouvoir répondre à des questions (requêtes) posées par des utilisateurs (grand public, professionnels, institutionnels, société civile, chercheurs, autres...).

Présentation générale :

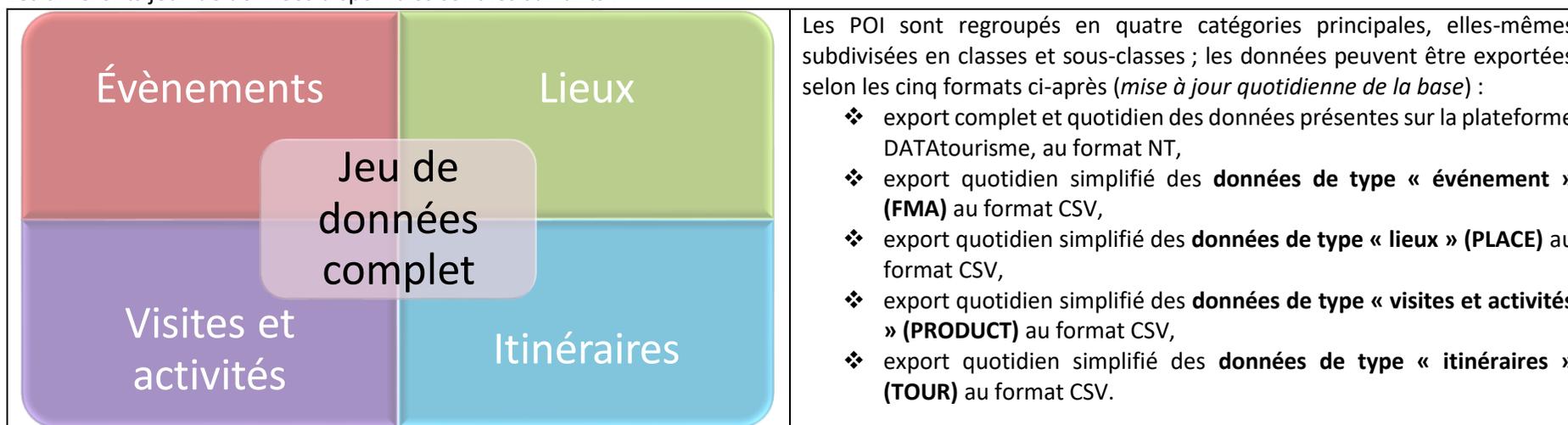
Voir : DATAtourisme, la base nationale des données du tourisme en Open Data - <https://www.data.gouv.fr/fr/datasets/datatourisme-la-base-nationale-des-donnees-du-tourisme-en-open-data/#>

DATAtourisme est un dispositif national de collecte et de diffusion en OpenData des données touristiques institutionnelles. Porté par l'Etat, il est né d'une coopération inédite avec les réseaux des offices de tourisme, des agences départementales et des comités régionaux du tourisme de l'ensemble du territoire national. Lauréat du Programme Investissements d'Avenir (PIA), ce dispositif participe à la boîte à outils gouvernementale pour décliner la stratégie touristique nationale telle que définie par le Comité Interministériel du Tourisme.

Directement connectée aux bases de données territoriales, la plateforme nationale DATAtourisme permet un accès plus large et simplifié aux données publiques d'information touristique, au moyen de :

- Une technologie capable d'agréger et mettre à jour quotidiennement plusieurs dizaines de jeux de données qualifiées,
- Une ontologie nationale permettant de retranscrire l'intégralité des données touristiques en un vocabulaire et un format uniformisé,
- Plusieurs services et une API permettant d'exploiter finement des données répondant à des besoins personnalisés,
- Une licence d'exploitation unique et ouverte (Licence ouverte 2.0).

Les différents jeux de données disponibles sont les suivants :



<p>❖ FETES ET MANIFESTATIONS (EntertainmentAndEvent) : manifestations, festivals, exposition, ou tout autre événement ayant un début et une fin.</p> <p>❖ FMA-exemple : SwimRun Gorges du Verdon, https://www.datatourisme.gouv.fr/ontology/core#EntertainmentAndEvent https://www.datatourisme.gouv.fr/ontology/core#PointOfInterest https://www.datatourisme.gouv.fr/ontology/core#SportsCompetition https://www.datatourisme.gouv.fr/ontology/core#SportsEvent http://schema.org/Event http://schema.org/SportsEvent , 43.773131, 6.205934, Base communale nautique et de plein air Quartier Margaridon, 83630 # Les Salles-sur-Verdon, 2020-11-01 <-> 2020-11-01, Office de Tourisme Intercommunal Lacs et Gorges du Verdon, Apidae Tourisme - source : Apidae Tourisme, 2020-04-01, # +33 6 49 42 35 90 # b.bouvier@alci4events.fr # http://www.alci4events.fr/ , , ""Plongez, l'effort ne fait que commencer....""</p>
<p>❖ LIEUX D'INTERET (PlaceOfInterest) : un lieu ayant un intérêt touristique (ex: un site naturel, un site culturel, un village, un restaurant, ...).</p> <p>❖ PLACE-exemple : Aups Tourisme, https://www.datatourisme.gouv.fr/ontology/core#LocalTouristOffice https://www.datatourisme.gouv.fr/ontology/core#PlaceOfInterest https://www.datatourisme.gouv.fr/ontology/core#Store https://www.datatourisme.gouv.fr/ontology/core#PointOfInterest https://www.datatourisme.gouv.fr/ontology/core#TouristInformationCenter http://schema.org/LocalBusiness http://schema.org/TouristInformationCenter , 43.627114, 6.224675, Place Martin Bidouré, 83630 # Aups, Office de Tourisme Intercommunal Lacs et Gorges du Verdon, Apidae Tourisme - source : Apidae Tourisme, 2020-04-10, # +33 4 94 84 00 69 # aups.tourisme@cclgv.fr # http://www.aups-tourisme.com/ <> http://www.lacs-gorges-verdon.fr/ , Accueil Vélo # France Vélo Tourisme, "Vivez Grandeur Nature."</p>
<p>❖ PRODUIT (Product) : un objet touristique qui peut se consommer (ex: une chambre d'hôtel, une pratique d'activité, une visite guidée, ...).</p> <p>❖ PRODUCT-exemple : Dragui Cycles, https://www.datatourisme.gouv.fr/ontology/core#PlaceOfInterest https://www.datatourisme.gouv.fr/ontology/core#Rental https://www.datatourisme.gouv.fr/ontology/core#Store https://www.datatourisme.gouv.fr/ontology/core#PointOfInterest https://www.datatourisme.gouv.fr/ontology/core#Product https://www.datatourisme.gouv.fr/ontology/core#EquipmentRental https://www.datatourisme.gouv.fr/ontology/core#ServiceProvider https://www.datatourisme.gouv.fr/ontology/core#EquipmentRentalShop http://schema.org/LocalBusiness http://schema.org/Product , 43.535239, 6.457542, 245 avenue du 4 Septembre, 83300 # Draguignan, Office de tourisme intercommunal Dracénie Provence Verdon, Apidae Tourisme - source : Apidae Tourisme, 2020-03-24, # +33 4 94 67 08 91 # btcdraguignan@gmail.com #, Accueil Vélo # France Vélo Tourisme, "Location de VTT, vélos de route et vélos électriques."</p>
<p>❖ ITINERAIRE TOURISTIQUE (Tour) : un itinéraire touristique est un POI qui propose un itinéraire composé d'étapes formant un parcours.</p> <p>❖ TOUR-exemple : Les Crinières du Verdon, https://www.datatourisme.gouv.fr/ontology/core # PlaceOfInterest https://www.datatourisme.gouv.fr/ontology/core#Tour https://www.datatourisme.gouv.fr/ontology/core#PointOfInterest https://www.datatourisme.gouv.fr/ontology/core#SportsAndLeisurePlace https://www.datatourisme.gouv.fr/ontology/core#HorseTour https://www.datatourisme.gouv.fr/ontology/core#EquestrianCenter http://purl.org/ontology/olo/core#OrderedList http://schema.org/LocalBusiness , 43.732887 , 6.584925 , Lieu-dit Saint Pierre Ferme Saint Pierre , 83840 # Bargème, Office de tourisme intercommunal Dracénie Provence Verdon, Apidae Tourisme - source : Apidae Tourisme, 2020-01-13, # +33 6 13 38 14 24 # juliette.czajko@lescrinieresduverdon.fr # https://www.lescrinieresduverdon.fr/ , , "Au cœur du Parc Naturel Régional du Verdon à la ferme Saint Pierre, venez profiter des chevaux dans un décor de rêve.", https://data.datatourisme.gouv.fr/13/e463f7fe-ce46-39c5-b075-d389d342ae90</p>

NB : Les informations sont disponibles en français et en anglais. Pour découvrir l'ontologie DATAtourisme, le format national sémantique pour les données d'information touristique, visitez le site : <https://framagit.org/datatourisme/ontology/tree/master>

Le système doit gérer un grand nombre d'éléments d'information et le tableau ci-dessous présente la nature et le volume actuel des informations disponibles (une information pouvant être un nom, , un mot, un chiffre, une date, un code postal, une coordonnée GPS, un lien Internet, un petit texte descriptif, etc.) :

Types de points d'intérêt touristique (POI)	Evénements (FMA)	Lieux (PLACE)	Visites et activités (PRODUCT)	Itinéraires (TOUR)
Catégories d'information requises (pour certaines informations, plusieurs éléments de réponses peuvent être concaténés)	<i>Nom_du_POI, Categories_de_POI, Latitude, Longitude, Adresse_postale, Code_postal_et_commune, Periodes_regroupees, Createur_de_la_donnee, SIT_diffuseur, Date_de_mise_a_jour, Contacts_du_POI, Classements_du_POI, Description, URI_ID_du_POI</i>	<i>Nom_du_POI, Categories_de_POI, Latitude, Longitude, Adresse_postale, Code_postal_et_commune, Createur_de_la_donnee, SIT_diffuseur, Date_de_mise_a_jour, Contacts_du_POI, Classements_du_POI, Description, URI_ID_du_POI</i>	<i>Nom_du_POI, Categories_de_POI, Latitude, Longitude, Adresse_postale, Code_postal_et_commune, Createur_de_la_donnee, SIT_diffuseur, Date_de_mise_a_jour, Contacts_du_POI, Classements_du_POI, Description, URI_ID_du_POI</i>	<i>Nom_du_POI, Categories_de_POI, Latitude, Longitude, Adresse_postale, Code_postal_et_commune, Createur_de_la_donnee, SIT_diffuseur, Date_de_mise_a_jour, Contacts_du_POI, Classements_du_POI, Description, URI_ID_du_POI</i>
Nombre mini d'informations requises par POI	14	13	13	13
Nombre de POI	Environ 152.000	Environ 349.000	Environ 43.000	Environ 10.000
... soit un total cumulé d'environ 554.000 POI.				

Ce tableau appelle les commentaires suivants :

- Le nombre de POI recensés est très variable selon le type de POI ;
- Les POI « événements (FMA) » nécessitent une information supplémentaire relative à la période d'activité (début et fin) ;
- Pour décrire certaines « visites et activités » (PRODUCT), par exemple tels que les établissements d'hébergement offrant plusieurs activités et services, on peut compter 25 éléments d'information, voire plus, pour un seul POI ;
- On peut évaluer le contenu en éléments d'information de la base à environ 10 millions.

Pour évaluer la représentativité de ces chiffres, il est important de comparer cette base de données avec les recensements disponibles en termes d'offre et de ressources touristique en France (*qui sont incomplets et ne recourent pas nécessairement le classement suivi par l'ontologie de DATAtourisme*) :

a) Etablissements d'hébergements touristiques en 2020 en France métropolitaine (INSEE, 06/04/2020) : total = 29.487, dont					
Catégories d'établissements d'hébergement	Hôtels de tourisme	Campings	Résidences de tourisme et hébergements assimilés	Villages de vacances et maisons familiales	Auberges de jeunesse, centres internationaux de séjour, centres sportifs
Nombre	17 492	8 312	2 416	1 001	266

NB : ces établissements offrent souvent plusieurs prestations complémentaires à l'hébergement qui ne sont pas décomptées ici.

b) Etablissements de restauration, cafés-bars et assimilés (FATIH 2018 et INSEE https://www.insee.fr/fr/statistiques/3600814) : total = 177.280, dont							
Catégorie	Restaurants traditionnels	Restauration collective concédée	Cafés-Bars	Traiteurs	Cafétérias	Casinos	Thalassothérapie
Nombre	110.000	14.200	37.000	15.000	830	200	50

NB : ces établissements offrent souvent plusieurs prestations complémentaires à l'hébergement qui ne sont pas décomptées ici (vente à emporter, drive, etc.).

c) Monuments historiques en 2018 et musées de France ouverts au public au 31 décembre 2017 en France métropolitaine (INSEE, 07/02/2020)		
	Monuments historiques classés ou inscrits en 2018	Musées de France ouverts au public au 31 décembre 2017
Nombre	44.776	1.077

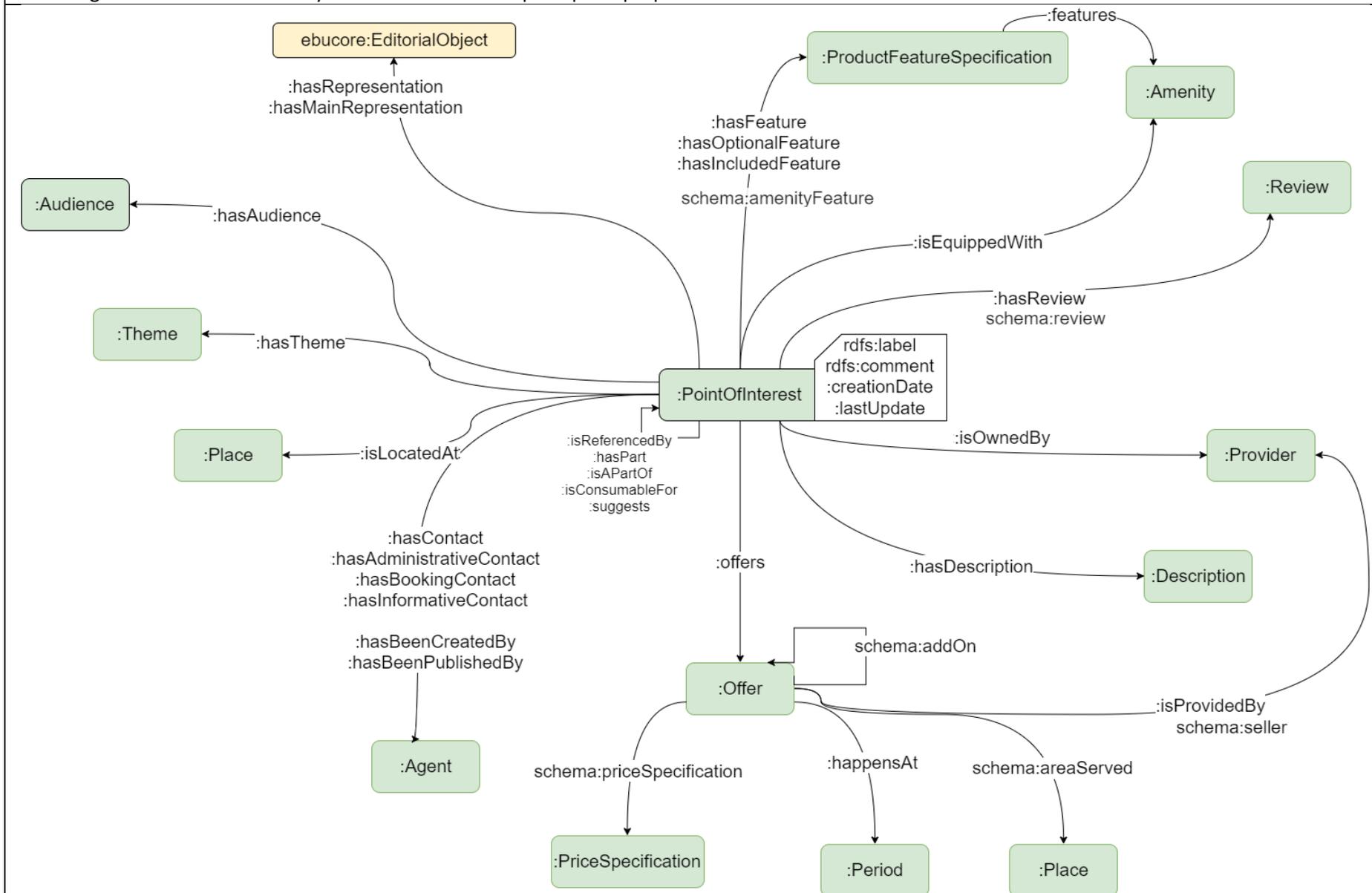
Citons également (entre autres) : [voir notamment : <https://www.culture.gouv.fr/Sites-thematiques/Developpement-culturel/Le-developpement-culturel-en-France/Culture-Monde-rural/Les-acteurs-et-les-outils-au-service-du-maillage-culturel-du-territoire/Les-partenaires/Protections-labels-et-appellations> ; <http://patrimoine.blog.pelerin.info/2019/01/04/combien-deglises-et-de-chapelles-en-france-en-2019/> ; voir également les Bases Mérimée - <http://www2.culture.gouv.fr/culture/inventai/patrimoine/> - et Joconde - <http://www2.culture.gouv.fr/documentation/joconde/fr/pres.htm>]

- 45.555 monuments historiques immobiliers et 260 000 objets mobiliers monuments historiques ;
- Plusieurs dizaines de milliers de châteaux et demeures ;
- 1.240 musées ;
- 1.500 parcs et jardins référencés ;
- 144 maisons remarquables ;
- 227 maisons des illustres ;
- 184 villes et pays d'art et d'histoire ;

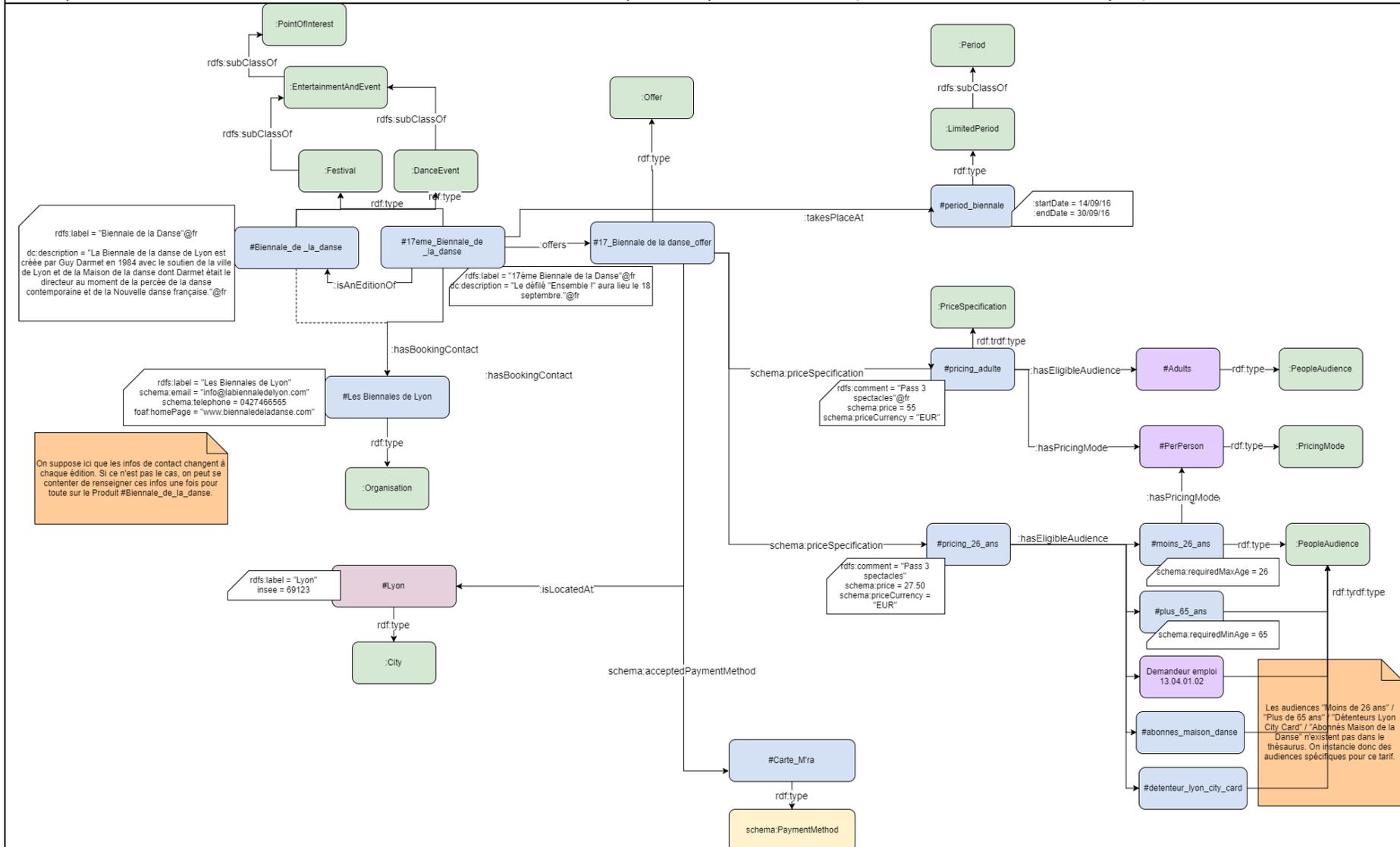
- Environ 2.300 édifices labellisés Patrimoine du XXème siècle ;
- 800 sites patrimoniaux remarquables ;
- 45 sites inscrits au Patrimoine mondial de l'Unesco, dont 39 culturels, 5 naturels et un mixte, plus 37 sur la liste indicative ;
- 38 sites Ramsar en France métropolitaine plus 12 outre-mer ;
- 17 sites remarquables classés ;
- 11 Parcs naturels nationaux ;
- 51 Parcs naturels régionaux (PNR) ;
- Des dizaines de milliers de manifestations culturelles, sportives, etc. ;
- Une centaine de festivals ;
- 29 itinéraires culturels ;
- Plus de 100.000 édifices religieux, dont 45.000 églises paroissiales et de nombreuses chapelles, ainsi que 2113 lieux de culte chrétiens réformés (protestants), 165 églises et chapelles orthodoxes, 448 synagogues, 1131 mosquées, 163 temples bouddhistes ;
- 59.517 Clochers de France ;
- 418 jardins remarquables labellisés ;
- 58 sites labellisés Sites remarquables du goût ;
- 74 grottes touristiques (ANECAT), 134 grottes ornées ou à gisement archéologique, 170 autres et 5 cavités remarquables ;
- 54 sentiers de grandes randonnées et 65 autres randonnées (<https://www.gr-infos.com/gr-fr.htm>) ;
- 154 parcs d'attractions et parcs à thèmes, ainsi que plus de 25 parcs de loisirs ;
- 4.135 piscines, 6.412 bassins de pratique de la natation et environ 25.000 plans d'eau ou étangs, dont 456 sous contrôle des agences de bassin (<https://www.eaufrance.fr/les-lacs-etangs-et-autres-plans-deau>) ;
- 9 routes historiques (<http://www.routes-historiques.com/>) ;
-

⇒ S'il est impossible de couvrir l'intégralité des ressources et des offres touristiques de la France, on arrive à un total d'environ 540/550.000 en additionnant toutes les ressources et offres ci-dessus ! il est tout à fait remarquable que cet ordre de grandeur soit très proche du cumul des POI identifiés dans DATAtourisme. Même si la nature de ces ressources et offres n'a pas pu être repointées entre les deux sources, on peut toutefois conclure que les ressources et offres présentes dans la base DATAtourisme correspondent à un volume « actualisé » des données ci-dessus dont certaines remontent à 2016, 2017 ou 2018. Leur nature et leur typologie devront toutefois être vérifiées ultérieurement.

Schéma général des relations du système d'information : principales propriétés et relations



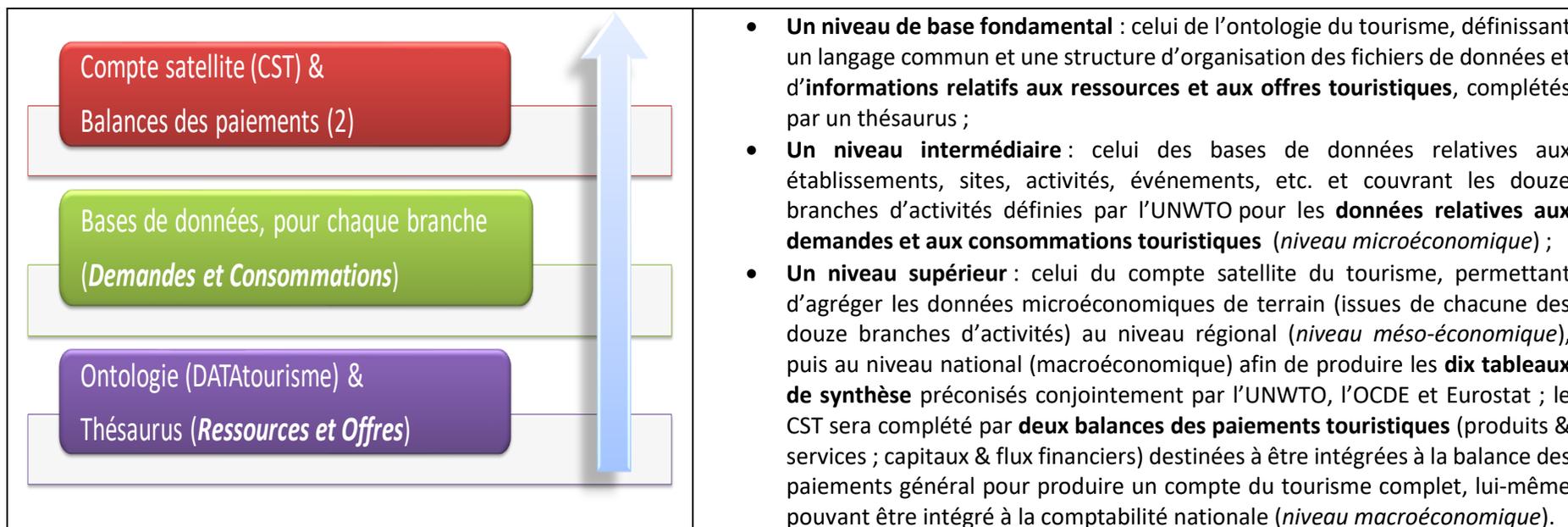
Exemple de structure de la base de données et des informations disponibles pour un festival (17^e biennale de la danse à Lyon) :



Vers un nouveau paradigme en matière d'observation du tourisme

Nous disposons donc désormais d'un outil « princeps », en l'occurrence l'ontologie et le thésaurus de la base de données relationnelle de DATAtourisme, lequel constitue le socle à partir duquel un nouveau modèle d'observatoire national du tourisme moderne peut être constitué. Dans la pratique, cet observatoire devrait avoir des relais régionaux afin de permettre une approche méso-économique, la plus adaptée pour un développement régional du tourisme et de ses activités.

L'avenir de l'observation, du management et d'un développement plus durable et résilient du tourisme peut donc désormais reposer sur ce niveau fondamental. Sur ce socle, vont alors pouvoir se (re)déployer deux niveaux constitutifs structurant un véritable système d'observation moderne du tourisme ainsi que le présente le schéma ci-dessous :



Conclusion

Le Cercle Stendhal se propose de mettre en œuvre sa boîte à outils EvalTo SETTI pour participer au développement et à la mise en œuvre de ce programme.