



APPLICATION DES TECHNOLOGIES 3.0 À LA DESCRIPTION DES ÉVÉNEMENTS CULTURELS



Publié en décembre 2016

Ce document est mis à disposition sous licence CC BY-SA 3.0 FR
(<https://creativecommons.org/licenses/by-sa/3.0/fr/>)

Table des matières

À propos de ce document	3
Résumé	4
Contexte	5
Les technologies sémantiques à la rescousse	7
Des modèles pour représenter des événements culturels	14
Comment accéder à des jeux de données événementielles	20
Un écosystème très riche d'applications pour présenter des agendas culturels	22
Avis, interactions sociales et recommandations	30
Conclusion	32
Remerciements	33
Glossaire	34
Références	36
Annexe A : description de l'événement « Liberté ! » à l'aide de schema.org	37

A PROPOS DE CE DOCUMENT

Ce document a été rédigé dans le cadre de la feuille de route stratégique « Métadonnées culturelles et transition web 3.0 » du Ministère de la culture et de la communication¹ par Raphaël Troncy d'Eurecom, Sophia Antipolis. Il fait suite aux échanges et réflexions menés au sein du groupe de travail "Appliquer les technologies 3.0 à la description des événements culturels" (GT n° 7), présidé par Rodolphe Bailly de la Philharmonie de Paris.

Il s'adresse en priorité aux personnels des établissements culturels qui gèrent les agendas, les réseaux sociaux institutionnels, les productions multimédias, les ressources documentaires en ligne et les sites web en général. Il a été validé dans la présente version le 30 novembre 2016.

¹ - Note publiée en janvier 2014 disponible à http://www.culturecommunication.gouv.fr/var/culture/storage/pub/feuille_de_route_metadonnees_culturelles_et_transition_web_3_0_janvier_2014/files/docs/all.pdf

RÉSUMÉ

Le Web est l'un des premiers vecteurs de diffusion et de valorisation de l'offre des institutions culturelles. Depuis quelques années, il détermine aussi les mécanismes de recommandation et, au final, les choix des sorties et visites culturelles. Cependant, les institutions n'ont aujourd'hui aucun moyen technique d'organiser entre elles, à grande échelle, l'interconnexion de leurs données décrivant leurs collections ou leur programmation, car leurs agendas en ligne fonctionnent en mode silo. Ce document montre, par l'exemple, comment l'utilisation de technologies sémantiques (ou dites orientées Web 3.0) permettra le décloisonnement par l'interconnexion de ces données. Il présente un panorama des modèles pour représenter ces données et il survole l'écosystème très riche des applications permettant de produire ou de consommer ces données, de visualiser et d'interagir avec des agendas culturels. Il offre enfin un certain nombre de préconisations quant à une évolution technique possible des systèmes d'information gérant les agendas culturels pour que ceux-ci bénéficient pleinement de la révolution des technologies dites 3.0. Il recommande de mettre en place un projet pilote, sur un territoire restreint mais culturellement très riche (e.g. le bassin de la Villette à Paris), qui reprendrait ces préconisations pour accroître la visibilité de l'offre culturelle de ce territoire.

Les bénéfices attendus sont multiples : pour le public, la possibilité d'effectuer des recherches thématiques transverses ou encore l'interconnexion des agendas avec les données de géolocalisation et de transport permettant un meilleur accès physique où les événements ont lieu; pour les centres de ressources culturelles, de nouveaux points d'accès pour leurs collections et leurs ressources, une meilleure visibilité dans les moteurs de recherche, le multilinguisme et la traduction favorisant le tourisme culturel; pour les acteurs culturels, un suivi précis de l'activité des créateurs (groupes musicaux, troupes de théâtre, danseurs...) tant au plan national qu'international. De manière générale, pour les professionnels de la culture, le recours à ces technologies et l'interconnexion de leurs agendas présentent un bénéfice certain en ressources humaines et moyens financiers sur le long terme.

CONTEXTE

L'offre en événements culturels est aujourd'hui exceptionnellement riche sur le plan national. Elle traverse l'ensemble des domaines de la culture et couvre largement le territoire. A titre d'exemple, pour les seuls secteurs de la musique, de la danse et du théâtre, on recense quelque 70 000 représentations pour la saison 2014-2015², programmées par les établissements publics et partenaires institutionnels du ministère de la Culture et de la Communication. Par ailleurs, la typologie de l'offre s'est étoffée au cours des dernières années (concerts, représentations théâtrales, expositions, salons, ateliers pédagogiques, rencontres, débats, visites-conférences, parcours ludiques intégrant des outils de réalité augmentée, visites tactiles pour non-voyants) tout en ciblant des publics très divers (publics de proximité, publics scolaires, touristes et visiteurs étrangers, publics empêchés).

Les institutions culturelles publient leurs agendas dans leurs sites web officiels et, depuis quelques années, les relayent via des espaces dédiés dans des réseaux sociaux. Pour éditer leurs agendas, elles utilisent des outils informatiques internes et des métadonnées qui sont cruciales pendant tout le cycle de vie de l'événement. Cependant, ces métadonnées reposent rarement sur des standards du web ce qui freine leurs réutilisations par des applications tierces. Dans une grande majorité des cas, ces métadonnées événementielles sont décorréliées de celles des objets et collections de l'institution. Enfin ces métadonnées sont également rarement pensées pour être archivées, ce qui donne lieu à de nouveaux processus d'édition et de copie, impliquant de fastidieuses ressaisies et un nombre croissant d'erreurs.

Côté diffusion, indépendamment des sites de sorties thématiques dédiés à des publics d'amateurs, les institutions n'ont aucun moyen technique d'organiser entre elles l'interconnexion de leurs données événementielles. Cette masse d'information, cloisonnée dans les sites officiels, fonctionne en mode silo. Force est de constater que la situation actuelle est centrée sur des logiques d'agendas institutionnels et non pas orientée sur des utilisateurs. Il va sans dire, pourtant, que les offres sont potentiellement complémentaires et que l'utilisateur final aurait un plus grand intérêt à explorer ces données une fois agrégées et combinées avec des données liées en provenance d'autres secteurs économiques tels que le transport ou le tourisme.

L'identification même des lieux, des établissements culturels et de leurs personnels clés souffre de problèmes d'interopérabilité. Prenons un exemple simple : la Maison de Victor Hugo située place des Vosges à Paris est décrite dans le répertoire des musées français MUSEOFILE par une notice documentaire identifiant ce lieu par M114³ et donnant, par exemple, la liste des oeuvres référencées dans la base Joconde. On retrouve cet identifiant – M114 – dans la base Patrimoine Numérique qui fournit des informations complémentaires telles que la liste des collections numérisées (lettres manuscrites et correspondances de l'écrivain) et les personnes contacts (directeur du musée, chargée des manuscrits/correspondance)⁴. La base de données EVE, qui gère les événements sur le portail Culture.fr, utilise un autre identifiant⁵ pour ce lieu (l'Hôtel Rohan-Guéméné) et pour rendre compte des expositions temporaires qu'il accueille. Ces différentes bases de données gagneraient à utiliser les mêmes identifiants de telle sorte que les informations minimales (e.g. les heures d'ouverture) ne soient pas recopiées avec le risque inhérent d'erreur et d'incompatibilité voire de contradiction, sans compter le gain de temps certain pour les professionnels à « propager » ces informations.

La promesse des technologies du web sémantique, encore dénommé web de données liées ou web 3.0, est justement d'offrir un cadre technologique standardisé et ouvert permettant d'identifier avec des URIs HTTP toute ressource sur le web et de décrire ces ressources à l'aide d'un modèle de données conceptuel (orienté graphes, par opposition à un modèle de données linéaire comme Dublin Core par exemple). Le démonstrateur JocondeLab⁶, résultat d'une expérimentation menée par le ministère de la Culture et de la Communication visant à démontrer l'apport des technologies du web sémantique, est ainsi un excellent exemple donnant accès à plus de 300 000 notices documentaires issues de la base de données Joconde, dans près de 14 langues grâce à un alignement des données avec Wikipédia.

2 - Mini-chiffres clés 2016, DEPS, Ministère de la culture et de la communication, <http://www.culturecommunication.gouv.fr/Politiques-ministerielles/Etudes-et-statistiques/Publications/Collections-d-ouvrages/Chiffres-cles-statistiques-de-la-culture/Chiffres-cles-2016>

3 - http://www.culture.gouv.fr/public/mistral/museo_fr?ACTION=CHERCHER&FIELD_98=REF&VALUE_98=M114

4 - <http://www.numerique.culture.fr/pub-fr/document.html?base=institution&id=FR-IN-M114>

5 - <http://www.culture.fr/Ressources/Labels/Maisons-des-illustres/Hotel-Rohan-Guemenee-Victor-Hugo>

6 - <http://jocondelab.iri-research.org>

Dans la suite de ce document, nous commençons par fournir un état des lieux de l'utilisation des technologies sémantiques pour la description des agendas culturels. Nous introduisons en particulier la notion de cartes de connaissances permettant aux moteurs de recherche du Web de fournir directement des réponses aux questions des utilisateurs. Nous décrivons ensuite les différents modèles, ou ontologies, permettant de représenter les événements culturels au sens large ainsi que les informations liées : l'offre en billetterie, l'offre de transport public pour s'y rendre, etc. Nous mettons aussi l'accent sur les discussions en cours pour stabiliser le vocabulaire permettant de décrire des événements culturels complexes. Nous proposons alors un inventaire des jeux de données événementielles ainsi que des applications, soient productrices, soient consommatrices de données décrivant des agendas culturels, en rappelant leur modèle d'affaire. Nous présentons un aperçu des techniques utilisées pour recommander aux utilisateurs des événements à venir et nous soulignons le rôle des réseaux sociaux, à la fois comme canal de diffusion de l'information, mais également comme lieu social propice aux interactions, avis et commentaires des participants. Nous concluons ce document en fournissant des préconisations destinées aux institutions culturelles pour accroître la visibilité de leur programmation culturelle sur le Web.

LES TECHNOLOGIES SÉMANTIQUES À LA RESCOURSE

Les technologies sémantiques sont spécialement conçues pour permettre l'interconnexion intelligente des données. Elles sont aujourd'hui mûres et constitueront, à court et moyen termes, un levier puissant pour la valorisation et la réutilisation des données de programmation culturelle.

Grâce à elles, les descriptions des événements culturels deviendront interoperables d'une institution à l'autre, et donc beaucoup plus facilement réutilisables. D'autre part, elles pourront être liées, non seulement aux référentiels institutionnels internes mais aussi à des référentiels externes du Web des données liées. Ainsi, les créateurs tels que les musiciens ou les troupes de théâtre pourraient être décrits à l'aide d'un identifiant ISNI⁷ ou VIAF⁸, une oeuvre musicale pourrait être enrichie grâce aux informations de MusicBrainz⁹ ou encore, les informations concernant le lieu d'un événement pourraient être complétées grâce à Wikipédia ou d'autres référentiels géographiques comme ceux de l'Insee par exemple.

Ce principe, consistant à avoir recours à des identifiants plutôt qu'à des chaînes de caractères pour décrire un artiste, un lieu, un thème, un objet ou une collection ou enfin un type d'événement, est la pierre angulaire de l'initiative du web des données liées. Il a été imaginé par Tim Berners Lee, l'inventeur du Web, dans un célèbre article intitulé « *The Semantic Web* » publié en 2001 dans la revue *Scientific American*¹⁰. Il a ensuite été popularisé le 16 mai 2012 dans un billet publié sur le blog de Google avec la devise : *things, not strings*¹¹ (« des choses et non des chaînes de caractères ») ! Google a alors révélé que son célèbre moteur de recherche sur le Web était plus performant grâce à un énorme graphe de connaissances (appelé *Knowledge Graph*) qui décrit précisément les points d'intérêts, les villes et les régions, les célébrités, les artistes et les sportifs, les produits, les films et les oeuvres artistiques, les bâtiments et même les étoiles, permettant ainsi au moteur de recherche de répondre directement aux questions posées par ses utilisateurs. Par exemple, la Figure 1 montre le résultat fourni par Google pour la requête « *qui sont les filles de Marie Curie ?* ». On observe qu'au-delà des traditionnels 10 liens bleus, la page de résultat s'est enrichie d'un panneau latéral droit contenant une carte de connaissance décrivant, ici, Marie Curie (quand elle est née ou ce qu'elle a découvert), et d'un panneau horizontal en haut de la page fournissant directement la réponse à la question, à savoir le nom des deux filles de Marie Curie. Selon Peter Norvig, directeur de la recherche à Google, ce graphe de connaissances revendique de décrire plus 600 millions d'objets à travers 20 milliards de faits ou de relations entre objets au mois d'avril 2016 (Norvig, 2016). Il est aussi au coeur de nouveaux services cherchant à anticiper les besoins d'information des utilisateurs à travers des notifications tels que Google Now¹².

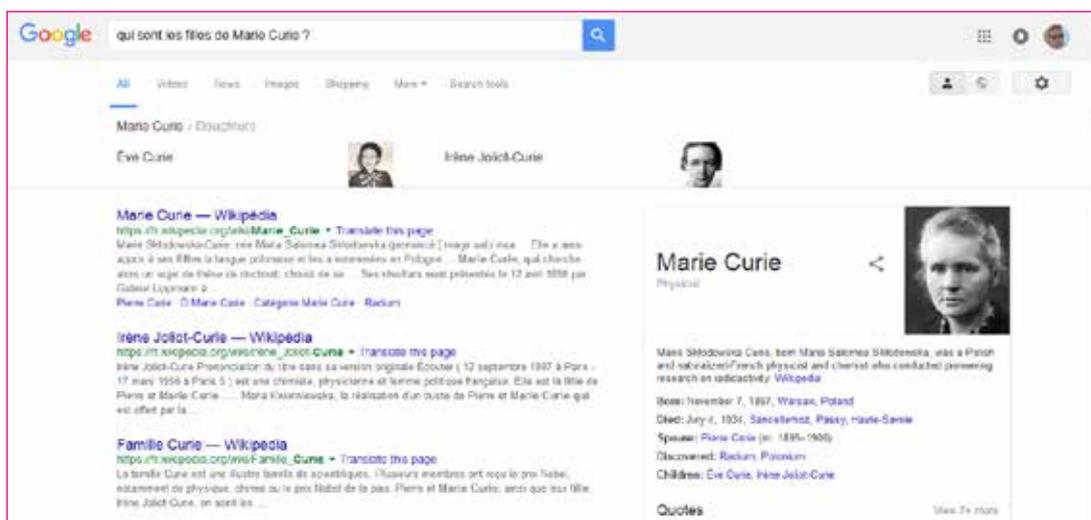


Figure 1 : A la question «qui sont les filles de Marie Curie ?», Google propose directement la réponse à la question (Eve Curie et Irène Joliot-Curie) en sus d'un panneau de connaissances décrivant Marie Curie

7 - International Standard Name Identifier (ISO 27729): <http://www.isni.org/>

8 - Virtual International Authority File: <https://viaf.org/>

9 - Encyclopédie de la musique libre: <https://musicbrainz.org/>

10 - <https://www.scientificamerican.com/article/the-semantic-web/>

11 - <https://googleblog.blogspot.fr/2012/05/introducing-knowledge-graph-things-not.html> et <https://www.youtube.com/watch?v=mmQl6VGuX-c>

12 - <http://time.com/google-now/>

L'idée consistant à vouloir répondre aux requêtes les plus fréquentes des utilisateurs directement dans les pages de résultats n'est pas propre à Google, et les autres moteurs de recherche sur le Web disposent tous de bases de connaissances, généralement organisées sous forme de graphes. En 2013, Microsoft Bing dévoilait Satori¹³, une technologie construite sur la base de connaissances ProBase¹⁴ qui permet elle aussi de générer des cartes de connaissances venant augmenter les résultats du moteur de recherche Bing. En Chine, le moteur de recherche Baidu a également annoncé en 2013 disposer d'une technologie similaire nommée ZhiXin, qui pourrait se traduire littéralement par «connaître le coeur (des choses)», et dont la base de connaissances est largement alimentée par l'encyclopédie chinoise Baidu Baike¹⁵. En Russie, le principal moteur de recherche Yandex offre à ces utilisateurs des cartes de connaissances grâce à la technologie Réponse par objets¹⁶ depuis 2015. En France, le moteur de recherche Qwant utilise principalement l'encyclopédie Wikipedia pour nourrir un Qnowledge Graph qui fournit des cartes de connaissances venant contextualisées les résultats de recherche¹⁷. Enfin, le moteur de recherche Yahoo doit être crédité pour avoir été le premier, dès 2008, à vouloir fournir une illustration des résultats de recherche à l'aide de cartes visuelles grâce à SearchMonkey¹⁸, une technologie qui venait lire les métadonnées structurées cachées dans les pages web et les restituait dans des gabarits personnalisables. Yahoo! Entity Graph est aujourd'hui encore une des plus grosses bases de connaissances.

Prenons un premier exemple : l'exposition *The Velvet Underground* s'est tenue à la Philharmonie de Paris du 30 mars au 21 août 2016¹⁹. La page web décrivant cet événement contient bien quelques données sémantiques cachées, produites par le système de gestion de contenu Drupal 7, ou automatiquement insérées pour faciliter le partage de l'événement sur les réseaux sociaux tels que Twitter ou Facebook avec un message pré-écrit.

Cependant, ces métadonnées doivent être améliorées pour créer une carte de connaissances de l'événement. Ce même événement est par ailleurs décrit sur les sites de vente ou de distribution de billets de spectacles tel que Ticketmaster²⁰. On trouve alors beaucoup plus de métadonnées sémantiques que les machines peuvent extraire et afficher ou simplement utiliser pour permettre l'achat d'un billet directement depuis une page de résultat ou un email. La Figure 2 montre le rendu sémantique de la page web Ticketmaster simulé par l'outil Structured Data Linter²¹ (à gauche) et l'outil Google Structured Data Tool²² (à droite). Ainsi, l'outil de Google précise qu'il a bien réussi à extraire la connaissance d'un événement, avec son nom, sa description, une image illustrative, sa durée et son adresse structurée. Il indique cependant que le marquage sémantique pourrait être complété par une description de l'artiste et de l'offre commerciale (prix et disponibilités des billets par catégories). L'outil Linter offre lui un rendu visuel tel qu'il pourrait apparaître dans un moteur de recherche.



Figure 2 : Rendu sous forme de carte de connaissances des métadonnées sémantiques incluses dans la page <http://www.ticketmaster.fr/fr/manifestation/the-velvet-underground-billet/idmanif/366701>. A gauche, le rendu simulé par l'outil Structured Data Linter. A droite, une validation sémantique proposée par l'outil Google Structured Data Testing Tool

13 - <http://blogs.bing.com/search/2013/03/21/understand-your-world-with-bing/>
 14 - <https://www.microsoft.com/en-us/research/project/probase/>
 15 - https://fr.wikipedia.org/wiki/Baidu_Baike
 16 - http://ru.wikipedia.org/wiki/Объектный_ответ
 17 - <https://blog.qwant.com/en/qwant-en/>
 18 - https://en.wikipedia.org/wiki/Yahoo!_SearchMonkey
 19 - <http://philharmoniedeparis.fr/en/activity/exposition/16455-velvet-underground> 20 - <http://www.ticketmaster.fr/fr/manifestation/the-velvet-underground-billet/idmanif/366701>
 21 - <http://linter.structured-data.org/>
 22 - <https://search.google.com/structured-data/testing-tool?hl=EN>

Prenons un second exemple : présenté dans le cadre du festival Jazz à la Villette 2016, *Pierre and the loup* est un événement musical pour enfants très attendu, présenté le samedi 3 septembre 2016, à la Philharmonie de Paris²³. Cet exemple montre que le site de la Philharmonie gagnerait également à améliorer la qualité de ses métadonnées sémantiques pour la description de ses concerts qui sont pour le moment trop peu développées. Pour le même spectacle, le site Ticketmaster²⁴ offre suffisamment de métadonnées pour générer des cartes sémantiques (Figure 3) tandis que le site de la FNAC²⁵ propose même des informations en temps réel concernant la disponibilité de billets (Figure 4). En l'occurrence, le site de la FNAC précise que cette disponibilité est épuisée via la valeur <http://schema.org/OutOfStock>. Toutefois, là encore, aucun site ne fournit des informations sémantiques pour en savoir plus sur les interprètes ou sur l'oeuvre de Serge Prokofiev qui a écrit ce conte musical en 1936 et dont ce spectacle est une adaptation.



Figure 3 : Rendu sous forme de carte de connaissances des métadonnées sémantiques incluses dans la page <http://www.ticketmaster.fr/fr/manifestation/pierre-and-the-loup-billet/idmanif/375542> selon l'outil Structured Data Linter

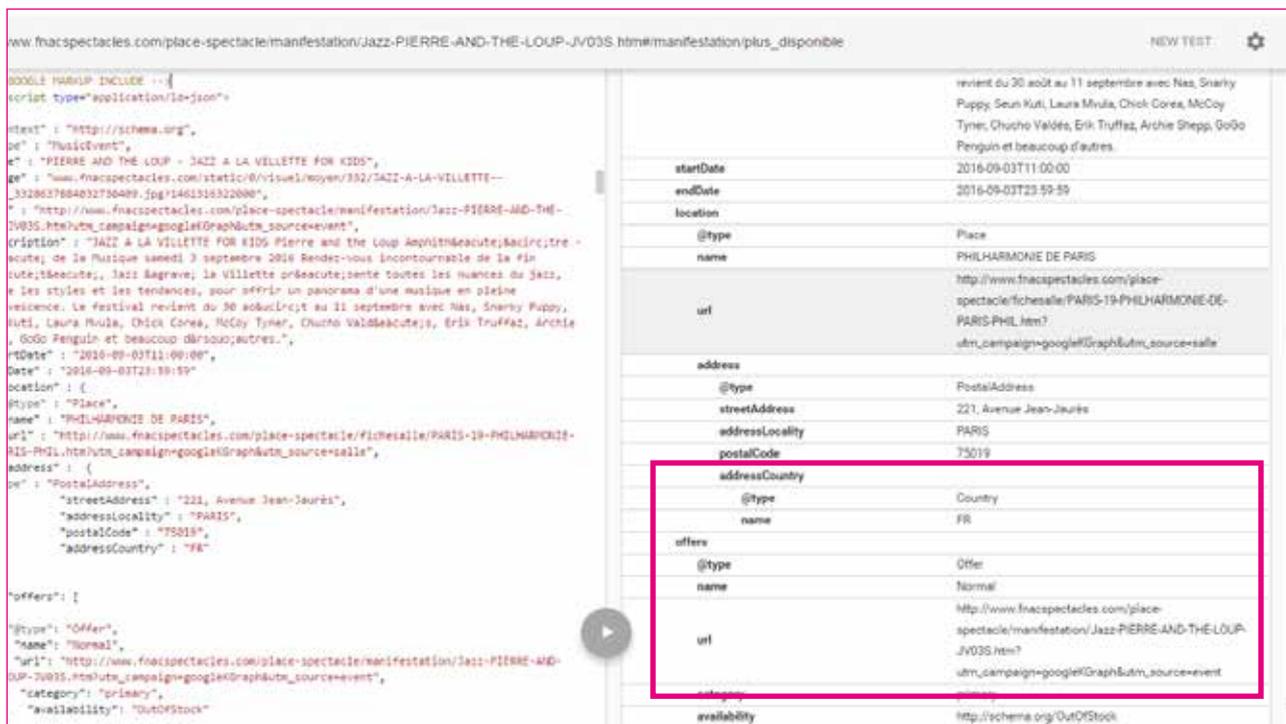


Figure 4 : Validation du marquage sémantique de la page <http://www.fnacspectacles.com/place-spectacle/manifestation/Jazz-PIERRE-AND-THE-LOUP-JV03S.htm#> à travers l'outil Google Structured Data. On y apprend par exemple que les billets sont épuisés pour cet événement

23 - <http://philharmoniedeparis.fr/en/activity/show-kids/17534-pierre-and-loup>

24 - <http://www.ticketmaster.fr/fr/manifestation/pierre-and-the-loup-billet/idmanif/375542>

25 - <http://www.fnacspectacles.com/place-spectacle/manifestation/Jazz-PIERRE-AND-THE-LOUP-JV03S.htm#>

Prenons un troisième et dernier exemple : les journées européennes du patrimoine sont désormais des manifestations nationales annuelles, qui ont lieu dans plus de 50 pays, sur le modèle des « Journées portes ouvertes des monuments historiques » créées en 1984 par le ministère de la culture français. En France et en Europe, les journées ont eu lieu les 17 et 18 septembre 2016²⁶ et le site officiel annonçait alors des milliers d'événements groupés en 13 catégories. Par exemple, la Cathédrale Sainte-Réparate à Nice accueillait le concert « Liberté ! » le samedi 17 septembre à 20:00 par l'ensemble vocal « Ristretto » et le « BBC Symphony Chorus » (Figure 5). Cet événement n'est cependant pas marqué sémantiquement et les moteurs de recherche ne pourront donc pas le proposer comme événements à venir dans ce lieu pourtant célèbre. Le site des Journées du Patrimoine qui s'appuie sur la solution logicielle d'Open Agenda comporte des données où un marquage sémantique est présent.



Figure 5 : Le concert «Liberté !» à la cathédrale Sainte-Réparate annoncée sur <http://journeesdupatrimoine.culturecommunication.gouv.fr/Programme?logid=1734794#14/43.6985/7.2656>



LocalBusiness	
@type	LocalBusiness
name	Cathédrale Sainte-Reparate
telephone	04 93 92 01 35
image	https://s3-media2.fl.yelpcdn.com/bphoto/fAsk7Uhrll1gmken0WmQ/ls.jpg
aggregateRating	
@type	AggregateRating
ratingValue	4.0
reviewCount	14
address	
@type	PostalAddress
streetAddress	3 place Rossetti
postalCode	06000
addressLocality	Nice
addressCountry	
@type	Country
name	FR
review	
@type	Review
datePublished	2010-10-11

Figure 6 : Carte de connaissances de Google pour la cathédrale Sainte-Réparate de Nice et programmation événementielle à venir (à gauche).
A droite, informations sémantiques extraites de la page Yelp <https://www.yelp.fr/biz/cath%C3%A9drale-sainte-reparate-nice-4?osq=Journées+du+patrimoine>

Le point commun entre les sites de billetterie tels que Ticketmaster ou la FNAC, ou les sites publiant des avis participatifs sur les commerces locaux ou les points d'intérêts est que tous intègrent des micro-données schema.org permettant aux robots d'indexation de mieux comprendre le sens des pages indexées. Schema.org est une initiative lancée conjointement par les moteurs de recherche Bing, Google et Yahoo! le 2 juin 2011, rejoint dès le 1er novembre 2011 par Yandex, qui a pour ambition de créer un ensemble commun de schémas pour structurer l'information sur le web et donc d'alimenter les bases de connaissances de ces moteurs de recherche.

La séparation du contenu de sa présentation a toujours été un des principes clés dans le développement du web, qui trouve désormais son essor sur les terminaux mobiles et où la présentation des informations doit nécessairement être adaptée aux petits écrans et aux modalités d'interaction tactile. L'ajout d'un marquage sémantique dans les pages n'est donc qu'une prolongation de ce principe. Invisible pour les utilisateurs qui lisent les pages, ces métadonnées peuvent être utilisées par les moteurs de recherche ou tout autre logiciel (réservation en ligne, comparaison de prix et d'avis, etc.) pour mieux répondre aux besoins d'informations. Schema.org puise son inspiration dans les travaux fondateurs de R. V. Guha, quand il travaillait pour Apple sur le projet Meta Content Framework (MCF)²⁷, et dans l'article visionnaire décrivant le web sémantique publié dans la revue Scientific American en 2001 (Berners-Lee, Hendler, & Lassila, 2001). Entre 1997 et 2004, le W3C a produit de nombreuses spécifications pour implémenter cette vision (RDF, RDFS, OWL, SKOS, SPARQL) tandis que de nombreux petits vocabulaires étaient proposés, souvent de manière ad-hoc, par des communautés, pour décrire des personnes et leurs relations, des adresses, des opinions ou encore des recettes de cuisine. Cette multitude de vocabulaires, utilisant des syntaxes différentes, parfois en compétition en ce qui concerne ce qu'ils décrivaient, ne facilitait pas leur adoption par les utilisateurs ou les moteurs de recherche. En 2011, les principaux moteurs de recherche se sont donc alliés pour proposer un seul vocabulaire centralisé, répondant aux questions les plus fréquentes des utilisateurs. Schema.org est ce vocabulaire qui a été à l'origine conçu pour répondre à des attentes essentiellement commerciales mais qui s'est et continue de s'adapter à d'autres domaines.

Aujourd'hui, schema.org est un vocabulaire de près de 600 classes, 850 propriétés et 115 valeurs énumérées, avec un mécanisme d'extension pour des communautés spécifiques telles que les bibliothèques et les musées²⁸, le domaine médical²⁹, ou encore le tourisme. Il est utilisé par des millions de site web, et sur un échantillon de 10 milliards de pages web indexées par Google, il a été rapporté que 31,3% d'entre elles contenaient du marquage sémantique utilisant schema.org (Guha, Brickley, & Macbeth, 2016). Si les moteurs de recherche sont les premiers consommateurs de ces métadonnées, ce ne sont pas les seuls : Siri, l'assistant intelligent d'Apple, utilise ces informations pour répondre aux questions. Pinterest, le réseau social permettant de partager ses centres d'intérêts sous forme d'albums de photographies glanées sur le web, utilise ces informations pour enrichir sémantiquement les objets de type recette de cuisine, film, produit ou événement qui sont épinglés. Parmi les sites les plus importants annonçant des événements culturels ou sociaux, <http://www.wherevent.com/>, <https://www.meetup.com/> ou <http://eventful.com/> ont été les premiers à adopter schema.org, et des sites tels que <http://news.bandsintown.com/> ou <http://www.ticketmaster.com/> ont vu leurs ventes de billets sensiblement augmenter depuis qu'ils ont adopté schema.org pour décrire les événements à venir (Guha, Brickley, & Macbeth, 2016).

Si schema.org est le vocabulaire recommandé par les principaux moteurs de recherche, il n'est pas le seul, et les administrateurs de site web sont aussi encouragés à utiliser d'autres vocabulaires sémantiques, en particulier ceux promus par les réseaux sociaux tels que Twitter ou Facebook. Ainsi, Twitter a lancé en juin 2012 le concept des cartes Twitter qui permet de créer des tweets multimédia enrichis³⁰. Pour mettre en place ces cartes, il suffit d'ajouter des métadonnées sémantiques particulières dans une page web, par exemple, celle annonçant un événement culturel à venir. Le fait d'inclure le lien de cette page dans un tweet créera alors automatiquement un tweet multimédia enrichi. Par exemple, L'Horizon est un collectif d'artistes du domaine plastique et du domaine vivant ayant pour objectif de produire des œuvres où les influences se croisent. La pièce de théâtre « Le Square, de Marguerite Duras » y a été présentée en mars dernier³¹, et la page annonçant ce spectacle inclut en particulier les métadonnées suivantes, où les attributs « *twitter:XXX* » sont à noter :

27 - https://en.wikipedia.org/wiki/Meta_Content_Framework

28 - <http://bib.schema.org/>

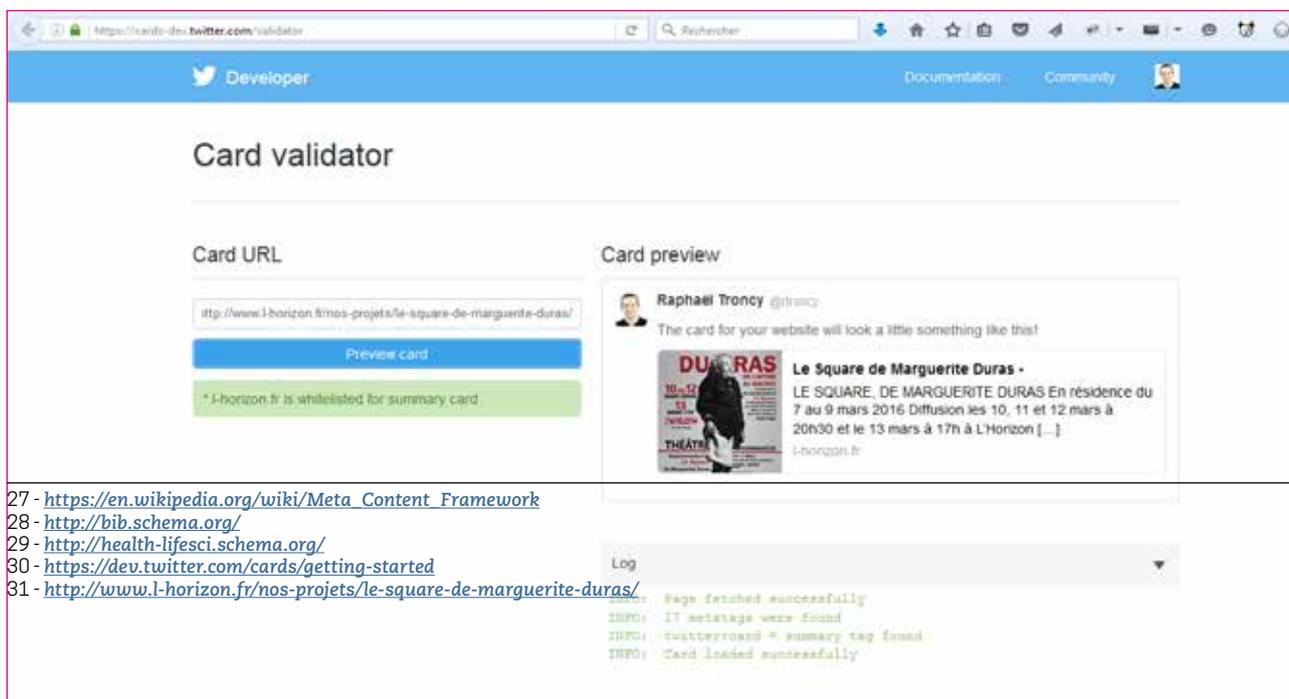
29 - <http://health-lifesci.schema.org/>

30 - <https://dev.twitter.com/cards/getting-started>

31 - <http://www.l-horizon.fr/nos-projets/le-square-de-marguerite-duras/>

```
<meta name="twitter:card" content="summary" />
<meta name="twitter:description" content="LE SQUARE, DE MARGUERITE DURAS En résidence du 7 au 9 mars 2016 Diffusion les 10, 11 et 12 mars à 20h30 et le 13 mars à 17h à L&rsquo;Horizon [&hellip;]" />
<meta name="twitter:title" content="Le Square de Marguerite Duras -" />
<meta name="twitter:site" content="@lHorizon_17" />
<meta name="twitter:image" content="http://www.l-horizon.fr/wp-content/uploads/2015/12/Affiche-web.jpg" />
<meta name="twitter:creator" content="@lHorizon_17" />
```

Ainsi, un tweet incluant le lien de cette page générera automatiquement le tweet multimédia montré dans la Figure 7 et l'administrateur du site aura également à sa disposition des informations analytiques quant à la viralité et le nombre de partages de cet événement sur le réseau social.



- 27 - https://en.wikipedia.org/wiki/Meta_Content_Framework
- 28 - <http://bib.schema.org/>
- 29 - <http://health-lifesci.schema.org/>
- 30 - <https://dev.twitter.com/cards/getting-started>
- 31 - <http://www.l-horizon.fr/nos-projets/le-square-de-marguerite-duras/>

Figure 7 : Simulation d'un tweet automatiquement généré grâce aux métadonnées sémantiques incluses dans la page <http://www.l-horizon.fr/nos-projets/le-square-de-marguerite-duras/>. Le simulateur est disponible à <https://cards-dev.twitter.com/validator>

Les institutions culturelles s'approprient de plus en plus les possibilités des réseaux sociaux. A noter cette initiative du Grand Palais pour promouvoir l'exposition consacrée à Hergé du 28 septembre 2016 au 17 janvier 2017 : un bot s'appuyant sur la technologie « Like a bird » a été mis en place afin de permettre aux twittos de se faire insulter avec le large éventail de jurons du capitaine Haddock, personnage haut en couleur de Tintin imaginé par Hergé.



Figure 8 : Compte Twitter @CaptHaddock_ alimenté par un robot à l'occasion de l'exposition Hergé qui a lieu au Grand Palais, <http://www.grandpalais.fr/fr/herge>

Facebook dispose d'un mécanisme similaire depuis 2010, et les administrateurs de site web sont aussi encouragés pour inclure des métadonnées permettant d'exister dans le graphe social nommé Open Graph³². Ainsi, cette même page web de l'Horizon annonçant cet événement culturel inclut aussi les métadonnées suivantes où les attributs « *og:XXX* » sont à noter :

```
<meta property="og:locale" content="fr_FR" />
<meta property="og:type" content="article" />
<meta property="og:title" content="Le Square de Marguerite Duras -" />
<meta property="og:description" content="LE SQUARE, DE MARGUERITE DURAS En résidence du 7 au 9 mars 2016 Diffusion les 10, 11 et 12 mars à 20h30 et le 13 mars à 17h à L&rsquo;Horizon &hellip;" />
<meta property="og:url" content="http://www.l-horizon.fr/nos-projets/le-square-de-marguerite-duras/" />
<meta property="og:site_name" content="L&#039;Horizon recherches et créations" />
<meta property="article:publisher" content="https://www.facebook.com/lhorizon17" />
<meta property="og:image" content="http://www.l-horizon.fr/wp-content/uploads/2015/12/Affiche-web.jpg"/>
<meta property="og:image" content="http://www.l-horizon.fr/wp-content/uploads/2015/12/affiche-le-square.jpg" />
```

Les statuts Facebook incluant ce lien seront automatiquement enrichis par une carte multimédia telle que montrée dans la Figure 8.

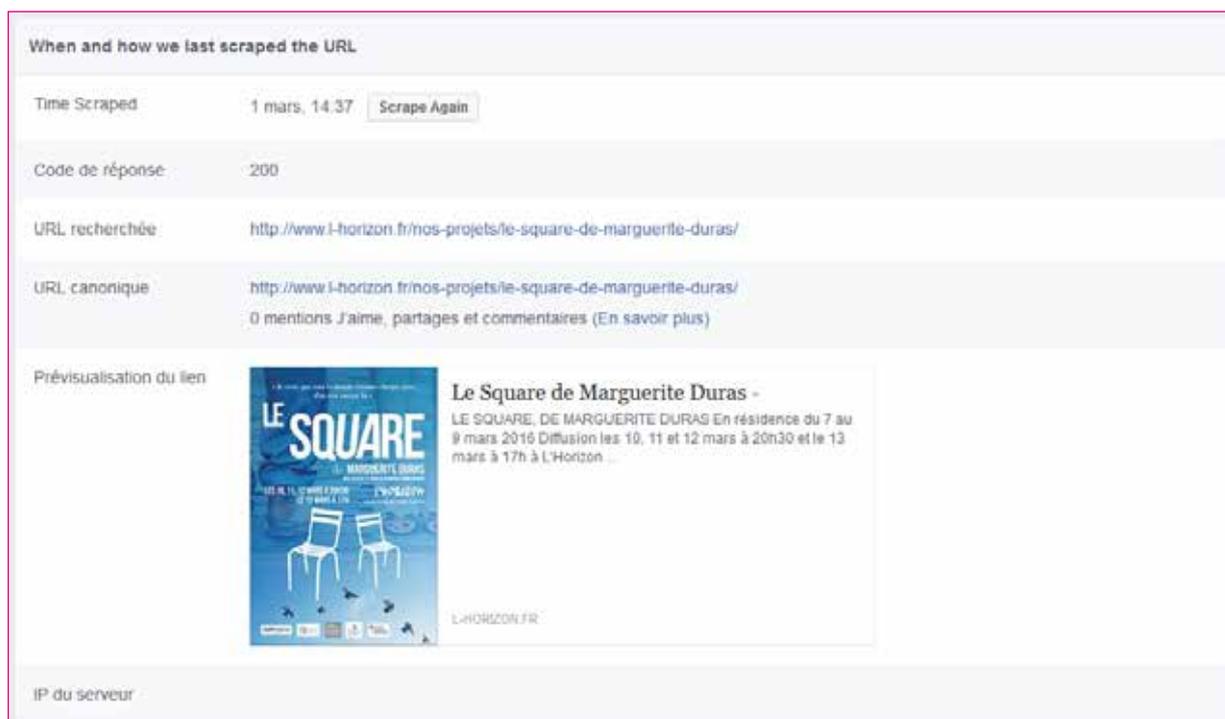


Figure 9 : Simulation d'un statut Facebook automatiquement généré grâce aux métadonnées sémantiques incluses dans la page <http://www.l-horizon.fr/nos-projets/le-square-de-marguerite-duras/>. Le simulateur est disponible à <https://developers.facebook.com/tools/debug/>

32 - <https://developers.facebook.com/docs/graph-api>

DES MODÈLES POUR REPRÉSENTER DES ÉVÉNEMENTS CULTURELS

Le terme événement est polysémique : il fait tout à la fois référence à des phénomènes passés (décrits dans des articles de presse ou expliqués par des historiens) et à des phénomènes planifiés dans le futur (notés dans un calendrier ou une programmation). Plusieurs formats et standards ont été proposés pour échanger des données structurées de calendrier (e.g. iCal).

Différentes ontologies, représentées en OWL ou RDFS et fournissant classes et propriétés pour décrire des événements et leurs relations ont été proposées (Table 1).

Ontologie	Namespace
CIDOC CRM	http://www.cidoc-crm.org/cidoc-crm/
Ontologie ABC	http://metadata.net/harmony/ABC/ABC.owl
Ontologie Event EO	http://purl.org/NET/c4dm/event.owl#
Ontologie SEM	http://semanticweb.cs.vu.nl/2009/04/event/
EventsML-G2	http://www.iptc.org/std/EventsML-G2/
DOLCE+DnS Ultralite	http://www.loa-cnr.it/ontologies/DUL.owl
F Event	http://events.semantic-multimedia.org/ontology/2008/12/15/model.owl
Ontologie OpenCYC	http://www.opencyc.org/

Table 1 : Ontologies pour représenter des événements

Bien que toutes les ontologies nommées ci-dessus contiennent des classes et des propriétés pour représenter le concept d'événement, elles ont été développées dans des buts différents. Ainsi, les ontologies CIDOC-CRM (Doerr, 2003) et ABC (Lagoze & Hunter, 2001) ont comme objectif de permettre l'interopérabilité des formats de métadonnées pour décrire des contenus multimédia possédés par les musées et les bibliothèques. Les événements visés sont donc historiques dans le sens large (e.g. guerre, naissance) ou liés aux objets décrits (e.g. changement de propriétaire, restauration, acquisition, prêt).

L'ontologie EO (Raimond, Abdallah, Sandler, & Giasson, 2007) a été développée pour être utilisée avec des ontologies musicales. Destinée principalement à la description d'interprétations artistiques, rien de spécifique au domaine musical n'a été cependant inclus. Elle est aujourd'hui l'ontologie la plus utilisée dans la communauté du web sémantique. L'ontologie SEM (van Hage, Malaisé, de Vries, Schreiber, & van Someren, 2009) permet, elle, de représenter des attaques maritimes dans des lieux parfois imprécis. EventsML-G2 est un modèle développé par l'IPTC³³ pour échanger des informations structurées décrivant les événements entre les fournisseurs d'information et leurs partenaires de la presse en ligne. Basé sur XML, ce standard permet de décrire des événements passés, impromptus et à venir.

DOLCE+DnS Ultralite (DUL) est une ontologie légère de haut niveau qui combine une axiomatisation simplifiée des concepts de DOLCE avec les patrons de représentation *Descriptions and Situations* (Gangemi & Mika, 2003). L'ontologie F Event est basée sur DUL dans la mesure où elle spécifie d'autres patrons pour représenter des relations (météorologique, participation, causale, corrélation) entre événements (Scherp, Franz, Saathoff, & Staab, 2009). OpenCYC est aussi une ontologie de haut niveau, mais suit une approche complètement opposée par rapport à DUL : plutôt que de fournir un ensemble très restreint de concepts, OpenCYC fournit des centaines de milliers de classes ayant pour but de modéliser « toute la réalité de l'homme ». Nous renvoyons le lecteur à (Troncy, Shaw, & Hardman, 2010) pour une comparaison exhaustive de ces modèles et de leur expressivité selon les dimensions agentives, temporelles, spatiales, causales ou compositionnelles.

33 - <https://iptc.org/>

L'ontologie LODE tire le meilleur parti des différents modèles étudiés ci-dessus puisqu'elle permet de représenter une description interopérable des aspects factuels d'un événement, ce qui peut se caractériser en terme des quatre Ws (en anglais What, Where, When et Who) : qu'est-ce qui s'est passé, où et quand cela s'est-il produit, qui était impliqué. Ces relations factuelles décrivant un événement ont comme objectif de représenter une réalité consensuelle et ne doivent donc pas être associées à une perspective ou une interprétation particulière.

Reprenons l'exemple du concert « Liberté ! » qui était donné à la cathédrale de Nice par l'ensemble vocal « Ristretto » et le « BBC Symphony Chorus » au mois de septembre à l'occasion des journées européennes du patrimoine. Cet événement peut se représenter avec l'ontologie LODE tel que représenté dans la Figure 7. Ce graphe indique qu'un objet identifié par 1735840 dans la base de données des Journées Européennes du Patrimoine est un événement de type Concert, qui aura lieu le 17 septembre 2016 à 20:30, dans un lieu géolocalisé par les coordonnées GPS (7.2758;43.6972) qui n'est autre que la Cathédrale de Nice, aussi identifiée par un réseau social tel que Foursquare qui fournit 2 conseils et avis et 68 photos. Enfin, ce graphe indique que ce concert sera interprété par l'artiste identifié sur Wikidata par Q4834982 qui n'est autre que le BBC Symphony Chorus, une chorale amatrice britannique basée à Londres fondée par la BBC en 1928.

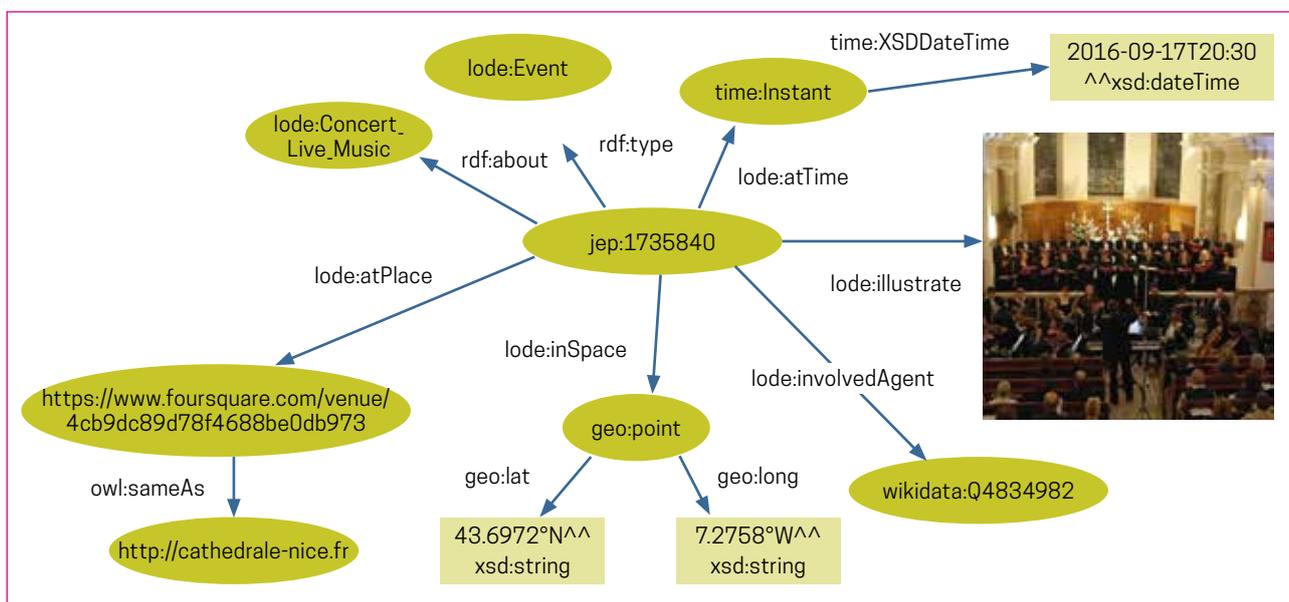


Figure 10 : L'événement «Liberté !» programmé pendant les Journées Européennes du Patrimoine représenté avec l'ontologie LODE

Ce même événement pourrait potentiellement être représenté en schema.org avec des métadonnées encore plus riches. Une description complète de cet événement est ainsi donnée en Annexe A, dans le format json-ld³⁴. Ces métadonnées décrivent d'abord l'événement « Liberté ! » (la date et le lieu du concert), mais aussi le fait que l'entrée soit libre ou donnent encore la liste détaillée des interprètes et du chef d'orchestre. L'événement englobant, à savoir les Journées Européennes du Patrimoine, peut par ailleurs être décrit lui aussi, et donner lieu à une autre carte de connaissances. De même, les oeuvres jouées telles que la « Figure Humaine », ou encore le compositeur (ici Francis Poulenc) peuvent être décrits pour venir enrichir ces métadonnées sémantiques. La Figure 10 montre ces quatre cartes de connaissances telles qu'elles pourraient être générées par les moteurs de recherche à partir des métadonnées fournies dans l'Annexe A.

34 - <http://json-ld.org/>

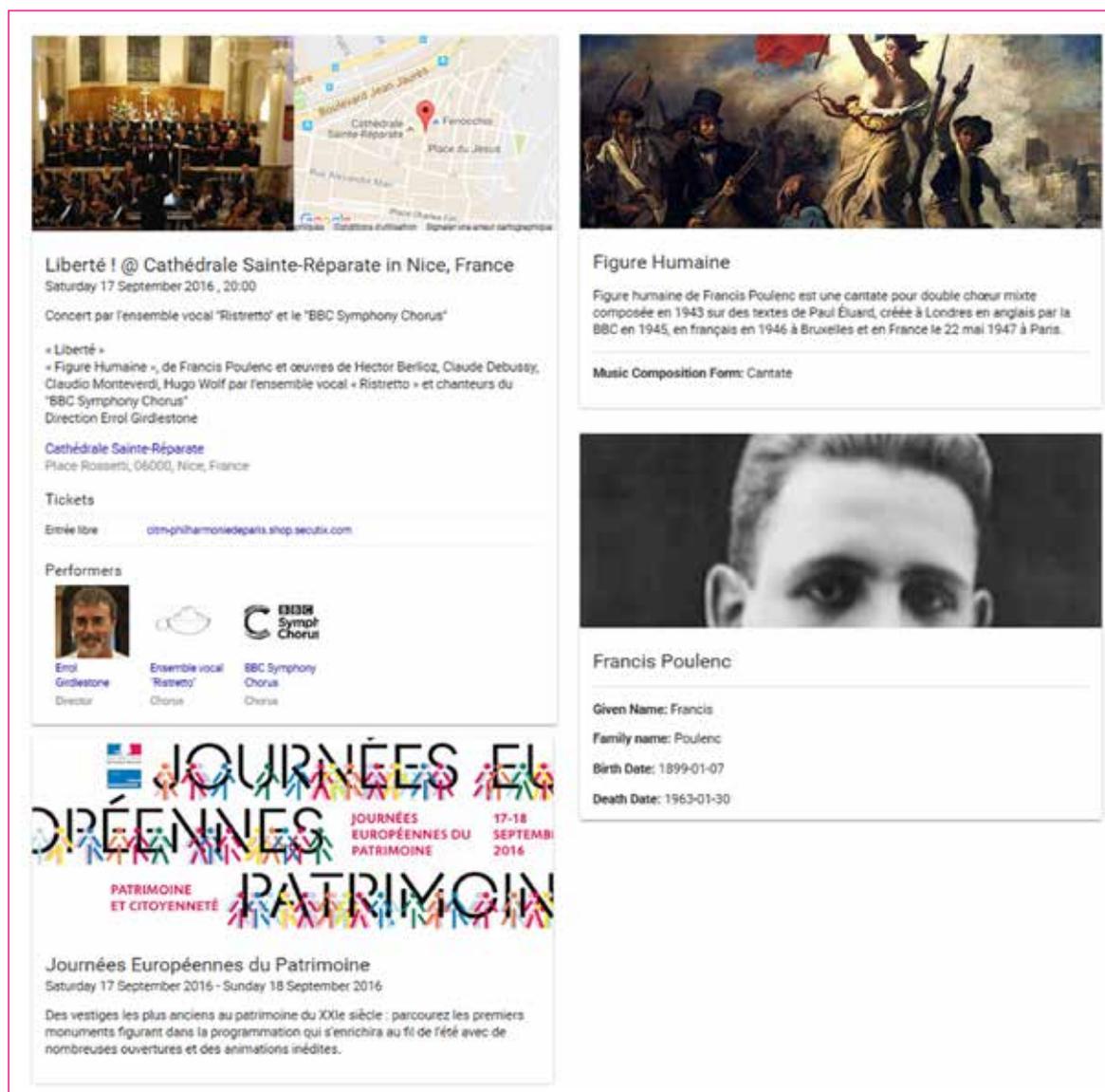


Figure 11 : Visualisation de 4 cartes de connaissances décrivant respectivement le concert « Liberté ! » et ses interprètes, l'événement global des Journées Européennes du Patrimoine, une des oeuvres interprétées, la « Figure Humaine » et son compositeur, Francis Poulenc. Ces cartes de connaissances sont visibles à <https://doremus-anr.github.io/schema-visualizer/#?id=57ada1d5bd966f0bda6b8a7d>

Ce rendu riche présenté dans la Figure 10 n'est toutefois pas encore disponible à travers les principaux moteurs de recherche et nous avons donc développé notre propre outil permettant de visualiser des métadonnées sémantiques qui utilisent le vocabulaire schema.org et qui décrivent des événements et des oeuvres culturels à <https://doremus-anr.github.io/schema-visualizer/>³⁵. Nous l'avons dit : le vocabulaire schema.org fournit près de 600 classes et 850 propriétés. Développer des cartes de connaissances pour tous ces types va donc prendre du temps. De plus, la modélisation des événements culturels n'est pas encore tout à fait stabilisée et un certain nombre de problèmes de modélisation sont toujours en cours de discussion. Nous listons ces problèmes dans les paragraphes ci-dessous.

Il est fréquent qu'un événement d'importance soit en fait composé d'une multitude de sous-événements dans un même lieu : c'est par exemple le cas d'un festival tel que le festival de théâtre et de spectacle vivant d'Avignon, le festival du film de Cannes ou encore les Francopholies de La Rochelle.

35 - Le code source est lui disponible à <https://github.com/DOREMUS-ANR/schema-visualizer/>

Un événement d'importance peut aussi être composé de nombreux sous-événements ayant lieu simultanément dans une multitude d'endroits : c'est par exemple le cas des Journées Européennes du Patrimoine ou encore de la Fête de la Musique ayant lieu tous les ans le 21 juin. Schema.org définit les propriétés `schema:subEvent` et `schema:superEvent` pour décrire des relations entre événements. Cependant, cette relation est aujourd'hui très faiblement utilisée par les sites Web et n'apparaît pas dans les cartes visuelles de connaissances. De même, il est fréquent qu'un événement se répète dans un même lieu mais à des dates différentes (ce sont par exemple les différentes représentations d'une pièce de théâtre), où même dans des lieux différents (ce sont par exemple les différents concerts d'une tournée d'un artiste, les différentes éditions des Jeux Olympiques ou encore, les différents matchs et événements spécifiques ayant lieu dans les fan zones pendant l'EURO). Jusqu'à présent, le point de vue de schema.org était de les considérer comme autant d'événements distincts et indépendants niant ainsi leur récurrence. Très récemment, la notion de `schema:EventSeries` a été proposée pour pallier ce problème mais la modélisation d'une série d'événements ne fait toujours pas consensus. Le lieu d'un événement doit ici s'entendre au sens large et inclut donc la notion de lieu virtuel ou de lieu en ligne. Ainsi, il est fréquent de voir des événements culturels retransmis en direct sur des sites dédiés, tels que <http://live.philharmoniedeparis.fr/>, ou sur des plateformes sociales de partage de vidéos. Là encore, Schema.org a prévu la notion de `schema:BroadcastEvent` pour décrire dans quelle condition un événement physique peut être retransmis en ligne.

Décrire les heures d'ouverture d'un lieu, en particulier quand celui-ci accueille un événement culturel, est primordial pour le consommateur de données. Cependant, décrire des heures d'ouverture, en toute généralité est complexe. Doit-on mieux spécifier des règles générales (par exemple, ouvert tous les jours de la semaine entre 9:00 et 18:00) suivies d'exceptions (par exemple, sauf les jours fériés, les derniers vendredi du mois, etc.) ? Ou, doit-on spécifier directement la liste complète sur l'année entière des jours et heures d'ouverture pour ne pas avoir à interpréter des règles et leurs exceptions ? Schema.org définit la propriété `schema:openingHours` pour décrire des règles simples ainsi que la classe `schema:OpeningHoursSpecification` pour décrire des ouvertures exceptionnelles mais il n'est toujours pas possible de décrire formellement quand des lieux sont exceptionnellement fermés.

La relation entre un événement culturel et une oeuvre, pour laquelle les institutions publiques disposent souvent de métadonnées riches, est souvent absente. Ainsi, aucune propriété ne fait le lien entre la description du concert « Liberté ! » et l'oeuvre la « Figure Humaine » de Francis Poulenc, ce qui donne lieu à différentes cartes de connaissances qui peuvent être juxtaposées (Figure 9). De même, le concert « Pierre and the Loup » décrit dans la Figure 3 donnera lieu à plusieurs cartes de connaissances montrées dans la Figure 11. Cette relation entre une oeuvre musicale et ses interprétations est au coeur des travaux du projet DOREMUS³⁶ financé par l'ANR qui publie une ontologie basée sur FRBRoo et des jeux de données interconnectés décrivant des oeuvres musicales, des CDs, des partitions et des concerts à partir de métadonnées fournies par trois grandes institutions culturelles : la Bibliothèque Nationale de France, Radio France et la Philharmonie de Paris.

36 - <https://github.com/DOREMUS-ANR/schema-visualizer/>

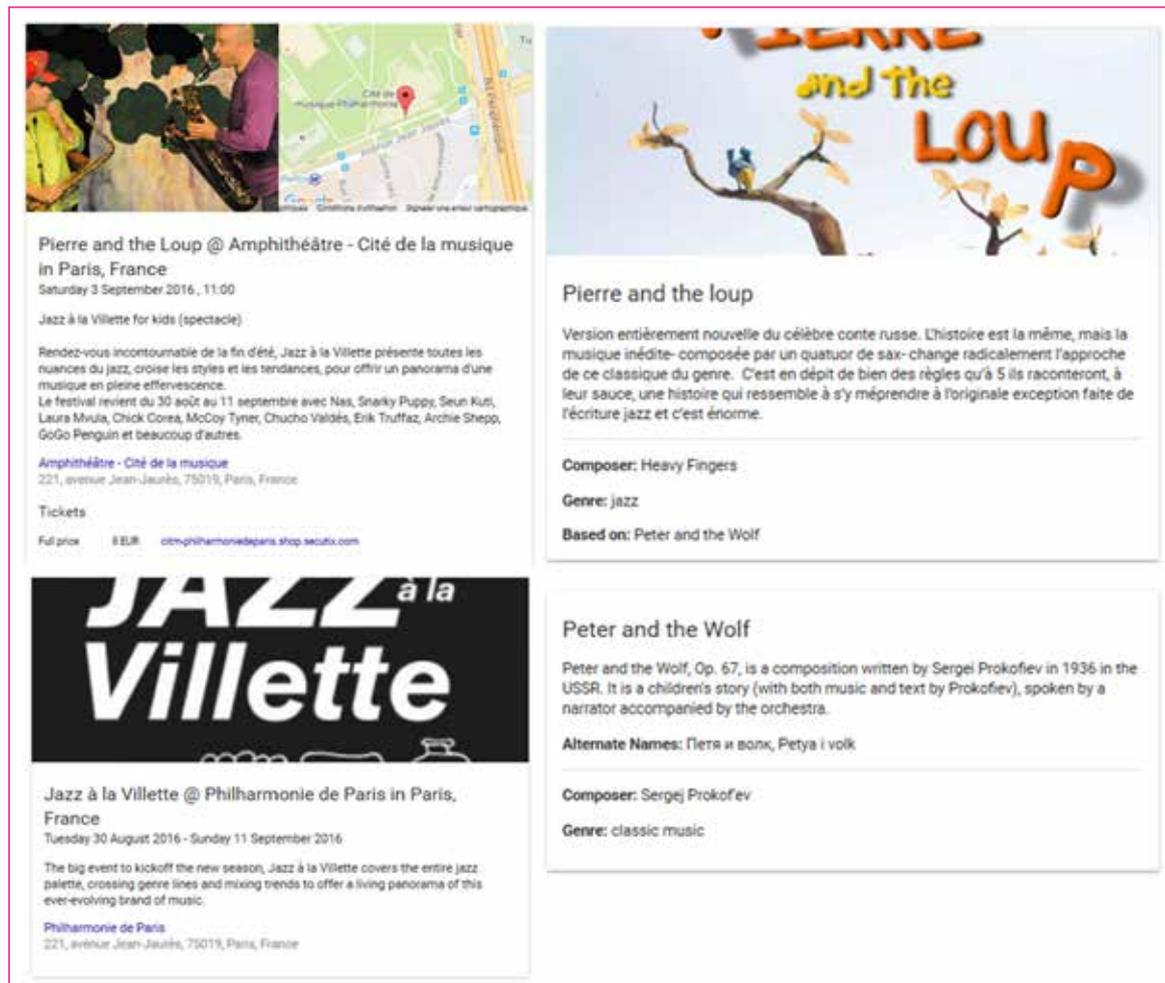


Figure 12 : Visualisation de 4 cartes de connaissances décrivant respectivement le concert « Pierre and the Loup », l'événement global du Jazz à la Villette, l'oeuvre interprétée par les « Heavy Fingers » et le conte original écrit par Sergei Prokofiev. Ces cartes de connaissances sont visibles à <https://doremus-anr.github.io/schema-visualizer/#?id=57ad9cc4c2e-f16159e54e36d>

Les événements sont toujours catégorisés, ce qui fournit un moyen pratique à l'utilisateur de parcourir les événements publiés selon leur type. Cependant, ces listes de type d'événements diffèrent selon les sites web. Schema.org propose 18 sous-types d'événement³⁷ (e.g. théâtre, sport, art visuel, exposition, festival, événements pour enfants) mais il est convenu que cette liste ne pourra jamais être exhaustive. Différentes dimensions peuvent rentrer en jeu pour catégoriser un événement, telles que le public visé, un genre musical particulier ou un sport. Les promoteurs de l'application Sowprog utilisent une liste de 9 catégories complétées par 47 styles permettant de raffiner le type d'événement. Les promoteurs de la solution Open Agenda travaillent eux aussi à une telle liste de type d'événements³⁸ et schema.org a reconnu ce problème³⁹ depuis plus d'un an. Dans des travaux précédents, nous avons manuellement analysé les taxonomies proposées par différents sites tels que Facebook, Eventful, Upcoming, Zevents, LinkedIn, EventBrite, TicketMaster ainsi que les jeux de données encyclopédiques du nuage de données. Nous avons alors appliqué la technique du tri par cartes pour construire un thésaurus de catégories d'événements contenant des renvois à ces sources (Khrouf & Troncy, 2014).

37 - <https://schema.org/Event>

38 - Voir le document de travail à https://docs.google.com/spreadsheets/d/1CsitZIHldzY434gi_fgeHAzxx_uhRA6AN33NYDn2C0/edit?usp=sharing

39 - <https://github.com/schemaorg/schemaorg/issues/406>

Le thésaurus est représenté en SKOS et les termes sont définis dans un espace de noms à <http://data.linkedevents.org/category/>. A terme, on pourrait espérer qu'une telle liste devienne un vocabulaire contrôlé incontournable qui serait hébergé sur la plateforme GINCO⁴⁰ du Ministère de la Culture et de la Communication.

La gestion de la billetterie est un enjeu majeur pour les programmeurs événementiels. Reprenant les codes du e-commerce, Schema.org définit le concept `schema:Offer` qui peut être attaché à un événement pour matérialiser la vente de billets : le prix, la garantie ou encore la disponibilité. S'il est possible de déclarer qu'un événement est complet (`schema:SoldOut`), il n'est en revanche pas possible d'indiquer le nombre exact de billets restant à la vente, ni les offres récurrentes de dernière minute, telles que la vente de billets à 5 euros sans réservation ouverte aux pièces de théâtre de la Comédie Française une heure avant le début de la représentation⁴¹.

Enfin, il est souvent utile de croiser les données liées à une programmation événementielle culturelle avec celles décrivant l'offre publique de transport sur un territoire donné. Le format GTFS (*General Transit Feed Specification*) s'est imposé ces dernières années comme le standard permettant de communiquer des horaires de transports en commun et les informations géographiques associées telles que la topographie d'un réseau (emplacement des arrêts, tracé des lignes). Une représentation sémantique de ce format nommée LinkedGTFS⁴² a été récemment proposée et il y a fort à parier que des alignements avec schema.org voient le jour dans les prochains développements. De même, les données décrivant la disponibilité en temps réel de places de parking dans une ville sont aujourd'hui extrêmement fragmentées et généralement incomplètes dues à la multiplicité des acteurs : collectivités territoriales, opérateurs privés. Parkopedia⁴³ se veut être un agrégateur de ces données et revendique aujourd'hui de fournir les places de stationnement disponibles dans plus de 6000 villes à travers 75 pays. Il est intégré avec des acteurs majeurs tels que les systèmes de navigation embarqués par les constructeurs automobiles ou le fabricant d'appareils mobiles Apple à travers son service de cartes. Cependant, aucun vocabulaire standard n'a encore été proposé pour représenter ces informations.

40 - <http://data.culture.fr/thesaurus/>

41 - <http://www.comedie-francaise.fr/billetterie-reservation.php?id=190>

42 - <https://github.com/OpenTransport/linked-gtfs>

43 - <http://www.parkopedia.fr/>

COMMENT ACCÉDER À DES JEUX DE DONNÉES ÉVÉNEMENTIELLES

Les institutions culturelles ont l'habitude de publier leurs agendas dans leurs sites web officiels. Elles utilisent, pour cela, des formulaires permettant de renseigner toutes les métadonnées décrivant ces événements : le lieu avec son adresse, sa géolocalisation et son accessibilité en transport en commun, la date et l'heure, les artistes, une catégorie, une description textuelle et des éléments visuels, le prix des billets, etc. Ces métadonnées sont parfois mises à disposition pour être réutilisées par des applications tierces venant présenter la programmation culturelle d'un lieu ou d'un territoire pendant toute une saison. Cette mise à disposition des données peut se faire à travers un export à instant donné sous forme de fichiers informatiques dans des formats structurés (le plus souvent, XML et CSV) ou à travers des APIs qui permettent, elles, d'être dans une logique de flux continu. Ces jeux de données événementielles, quand ils sont disponibles, sont toutefois disséminés à travers de multiples portails de données, dans des formats différents et attachés à des licences variées même si la licence ouverte Etalab⁴⁴ s'impose peu à peu.

EVE⁴⁵ est la base de données du ministère de la Culture et de la Communication qui nourrit le portail <http://www.culture.fr/>. Celui-ci encourage les organismes culturels à référencer toute manifestation ayant un rapport avec le domaine culturel en France (festival, concert, exposition, conférence, spectacle de danse, théâtre, atelier, visite, etc.). S'il n'est pas possible d'avoir accès aux données d'EVE en continu, l'agenda de l'offre culturelle est exporté en fin de saison et archivé sur le portail de données public data.gouv.fr mis en place par la mission Etalab. On peut ainsi retrouver l'agenda culturel pour les années 2012 à 2015⁴⁶.

Outre l'agenda Culture, le ministère de la Culture et de la Communication met à disposition d'autres jeux de données sur le portail data.gouv.fr, pour des événements spécifiques tels que les Journées Européennes du Patrimoine⁴⁷, la Fête de la Musique⁴⁸ ou la Nuit des Musées⁴⁹. Cependant, ces jeux de données ne deviennent disponibles qu'après que les événements ont lieu confirmant une logique d'archivage de la métadonnée, qui plus est, dans des formats et des structures de données qui changent tout le temps.

Le ministère de la Culture et de la Communication n'est pas le seul producteur de données événementielles et d'autres organismes utilisent également le portail data.gouv.fr pour archiver leur programmation culturelle à l'issue d'une saison. Ainsi, le Musée du Louvre a mis à disposition cet agenda pour la seule saison 2013-2014⁵⁰. On remarque, cependant, que ces données ne sont pas mises à jour et que leur mise à disposition n'est pas encore inscrite dans une véritable automatisation du processus de publication de l'agenda culturel.

De très nombreuses régions et collectivités territoriales mettent également à disposition des données ouvertes à travers des portails locaux. Ceux-ci offrent généralement des données mises à jour très régulièrement. Ainsi, l'Office du Tourisme et des Congrès de Nice propose une mise à jour quotidienne des événements publics recensés sur le territoire de la Côte d'Azur via le portail de données de la métropole⁵¹ mais ces données sont souvent incomplètes : une adresse mais pas le nom du lieu accueillant l'événement, des horaires d'ouverture absente ou encore une description limitée à 10 mots. Dans la région du grand ouest, data.locale.fr fait figure d'exemple, proposant plus de 600 000 événements par an dans un jeu de données mis à jour quotidiennement⁵², également recopié sur le portail national data.gouv.fr⁵³. Cependant, là encore, chaque collectivité territoriale ou office du tourisme propose son propre modèle de données ad-hoc, sans véritable concertation, freinant ainsi la réutilisation de ces données par une application qui voudrait avoir comme ambition une couverture nationale, même si le secteur des données touristiques semble se consolider avec des modèles tels que APIDAE⁵⁴.

44 - <https://www.etalab.gouv.fr/licence-ouverte-open-licence>

45 - <http://www.culture.gouv.fr/nav/acceseve.htm>

46 - <https://www.data.gouv.fr/fr/datasets/agenda-de-l-offre-culturelle/>

47 - <https://www.data.gouv.fr/fr/datasets/journees-europeennes-du-patrimoine/>

48 - <https://www.data.gouv.fr/fr/datasets/programme-de-la-fete-de-la-musique-2015/>

49 - <https://www.data.gouv.fr/fr/datasets/nuit-des-musees/>

50 - <https://www.data.gouv.fr/fr/datasets/programmation-evenementielle-et-culturelle-du-musee-du-louvre-576199/>

51 - <http://opendata.nicecotedazur.org/data/dataset/liste-des-evenements-publics-recenses-par-l-office-du-tourisme>

52 - http://datainfocale.opendatasoft.com/explore/dataset/agenda_culturel/

53 - <https://www.data.gouv.fr/fr/datasets/donnees-evenementielles-infocale/>

54 - <http://apidae-tourisme.com/>

Philharmonie de Paris (Q3378907)
 巴黎音樂廳
 concert hall in Paris

Location properties
 Instance of: Concert hall

Maps
 A map showing the location of the Philharmonie de Paris in Paris, France, with zoom controls and a search bar.

External sites
 official website

External sources
 Emporis building: 288071
 Freebase: /m/0411x84
 MusicBrainz place: 0c37111b-0464-4389-b09a-96532e036162
 Structure ID: 20088040

External identifiers

Wikimedia projects
 Current language Wikipedia: en: Philharmonie de Paris
 Big Wikipedia: ca: Filharmonia de París, de: Philharmonie de Paris, es: Filarmónica de París, fr: Philharmonie de Paris, it: Filharmonia de Paris, ja: フランス国立交響楽団, zh: 巴黎愛樂樂團
 Wikimedia Commons: Category:Philharmonie de Paris
 Other Wikipedia: vi: Philharmonie de Paris

Location

Name	Description
Europe	continent in Earth's Northern Hemisphere
France	country in western Europe
Île-de-France	region in France where the French capital Paris is located
Paris	capital city of France
Philharmonie de Paris	concert hall in Paris

Other properties

significant event	Inauguration: Formal ceremony to mark the beginning of a major public leader's term of office, or official opening or beginning of an institution or structure. part in time: 2015-01-14 [A1]
architect	Jean Nouvel: French architect (b. 1945)

Related media
 Commons category: Philharmonie de Paris

Media thumbnails: A portrait of Jean Nouvel, a map of France, a view of the Philharmonie de Paris building, the Île-de-France logo, and a map of Europe.

Figure 13 : Visualisation des données décrivant l'entité Q3378907 identifiant la Philharmonie de Paris dans l'outil Reasonator, <https://tools.wmflabs.org/reasonator/?&q=3378907>

Les artistes sont souvent identifiés par un numéro unique géré par l'International Standard Name Identifier (ISNI⁵⁹). Cette base de données s'est constituée à partir de novembre 2011 sur le socle VIAF⁶⁰, un fichier d'autorité international virtuel permettant d'identifier les personnes et les collectivités, largement utilisé par les bibliothèques. Des mises à jour régulières et réciproques ont lieu entre les bases ISNI et VIAF qui ont toutes deux adoptées les technologies du web sémantique matérialisées par des liens actifs entre les jeux de données. Les jeux de données événementielles utilisent, eux, rarement ces liens et identifiants pour référencer les lieux et les artistes, privilégiant de simple chaînes de caractères et introduisant ainsi une certaine ambiguïté.

59 - <http://www.isni.org/>
 60 - <http://viaf.org/>

UN ÉCOSYSTÈME TRÈS RICHE D'APPLICATIONS POUR PRÉSENTER DES AGENDAS CULTURELS

On recense de nombreuses applications permettant de gérer ou de présenter des agendas culturels. Parmi les applications productrices de données événementielles, on trouve des agendas participatifs qui permettent de rentrer de la donnée via des formulaires simples tout en promettant une promotion maximale de l'événement par la rediffusion des données sur un nombre toujours plus grand de sites web : l'enjeu consiste ici à assurer le meilleur référencement possible des événements culturels et le modèle d'affaire de ces sociétés fonctionne généralement par abonnement pour ceux cherchant à promouvoir un événement ou une salle. Certaines applications proposent d'alimenter leur base de données d'événements à venir de manière automatique, soit en allant chercher l'information noyée dans les pages web à l'aide d'outils d'apprentissage et de fouille de données dédiés, soit en allant chercher l'information sur les principaux réseaux sociaux tels que Facebook. Souvent, ces applications offrent en retour un accès aux données ainsi collectées et permettent donc certaines formes de réutilisation.

Parmi les applications consommatrices de données événementielles, on trouve essentiellement des applications de recommandation dont l'objectif est de proposer à un utilisateur une sortie dans l'instant. Le modèle d'affaire de ces applications est adossé à la billetterie, une commission étant prélevée sur les ventes.

Dans la suite de ce chapitre, nous présentons les principales applications qui illustrent la richesse de cet écosystème. Nous décrivons également quelques projets de recherche, fortement orientés « technologies », qui offrent souvent une vision plus large de la réutilisation des données événementielles combinées avec d'autres secteurs d'activités. Nous mentionnons enfin des projets et applications qui ont cessé de fonctionner, par manque de financement, illustrant donc la difficulté de trouver un modèle d'affaire pérenne dans ce marché. La Table 2 propose une comparaison de ces différentes applications selon plusieurs critères : leur année de lancement, leur type, leur couverture géographique, les sources de données utilisées, le nombre d'événements décrits en base et les technologies utilisées.

	Année	Type	Couverture	Sources	Volume	API	Technologies
Applications productrices de données événementielles							
Open Agenda	2011	Commercial	France	Saisie manuelle collaborative	125 000 événements	Oui	modèle de données schema.org
Sowprog	2012	Commercial	France (80% Paris)	Saisie manuelle collaborative	30 000 événements	Oui	modèle de données ad-hoc / publication utilisant schema.org
Agenda Culturel	2005	Commercial	France	Saisie manuelle collaborative	150 000 événements	Oui	modèle de données ad-hoc
Mapado	2012	Commercial	France	Sites web des lieux culturels + Données ouvertes	inconnu	Non	Apprentissage et fouille de données
Evensi	2013	Commercial	Monde	Facebook	48 000 000 événements	Oui	modèle de données ad-hoc

	Année	Type	Couverture	Sources	Volume	API	Technologies
Applications consommatrices de données événementielles							
Toot Sweet	2015	Commercial	France (95% Paris)	Saisie manuelle collaborative + partenariat avec les acteurs de la billetterie	2000 sorties ciné + 200 événements / jour	Non	modèle de données ad-hoc
EventMedia	2009	Projet de recherche	Monde	Last.fm + Eventful + Flickr	110 000 événements	Oui	web sémantique / ontologie LODE
3cixty	2014	Projet de recherche	Milan, Nice, Londres	Evensi + Last.fm + Eventful + Office Tourisme + logique éditoriale	40 000 événements	Oui	web sémantique / schema.org
Mapado	2012	Commercial	France	Sites web des lieux culturels + Données ouvertes	inconnu	Non	Apprentissage et fouille de données
Evensi	2013	Commercial	Monde	Facebook	48 000 000 événements	Oui	modèle de données ad-hoc
Projet et applications qui ont cessé d'exister							
Open Data Events	2013	Open Source	Nantes	Saisie manuelle collaborative	N/A	Oui	modèle de données ad-hoc
Calendeev	2013	Commercial	France	Données ouvertes	N/A	Non	modèle de données ad-hoc

Table 2 : Comparaison de différentes applications productrices ou consommatrices de données événementielles

OpenAgenda⁶¹ est une plateforme participative permettant à des acteurs de créer, de gérer et de publier sous formes de données ouvertes des événements géolocalisés et des agendas (qui regroupent des événements). OpenAgenda fournit également sa technologie en marque blanche pour s'insérer dans n'importe quel site web. L'application, lancée en 2015, était précédemment nommée Cibul qui a commencé à fournir des données dès le mois de juillet 2011. A ce jour, OpenAgenda revendique de lister plus de 125 000 événements, et près de 6 000 agendas, ces données venant de près de 25 000 contributeurs. La Figure 13 montre, par exemple, l'agenda officiel des Journées Européennes du Patrimoine 2016 avec près de 400 événements prévus pour la seule région PACA. OpenAgenda compte parmi ses clients les régions Bretagne et Ile de France, le NUMA, le Conseil National du Numérique, la FrenchTech, Paris.fr ou encore le Ministère de la Culture et de la Communication.

61 - <https://openagenda.com/>

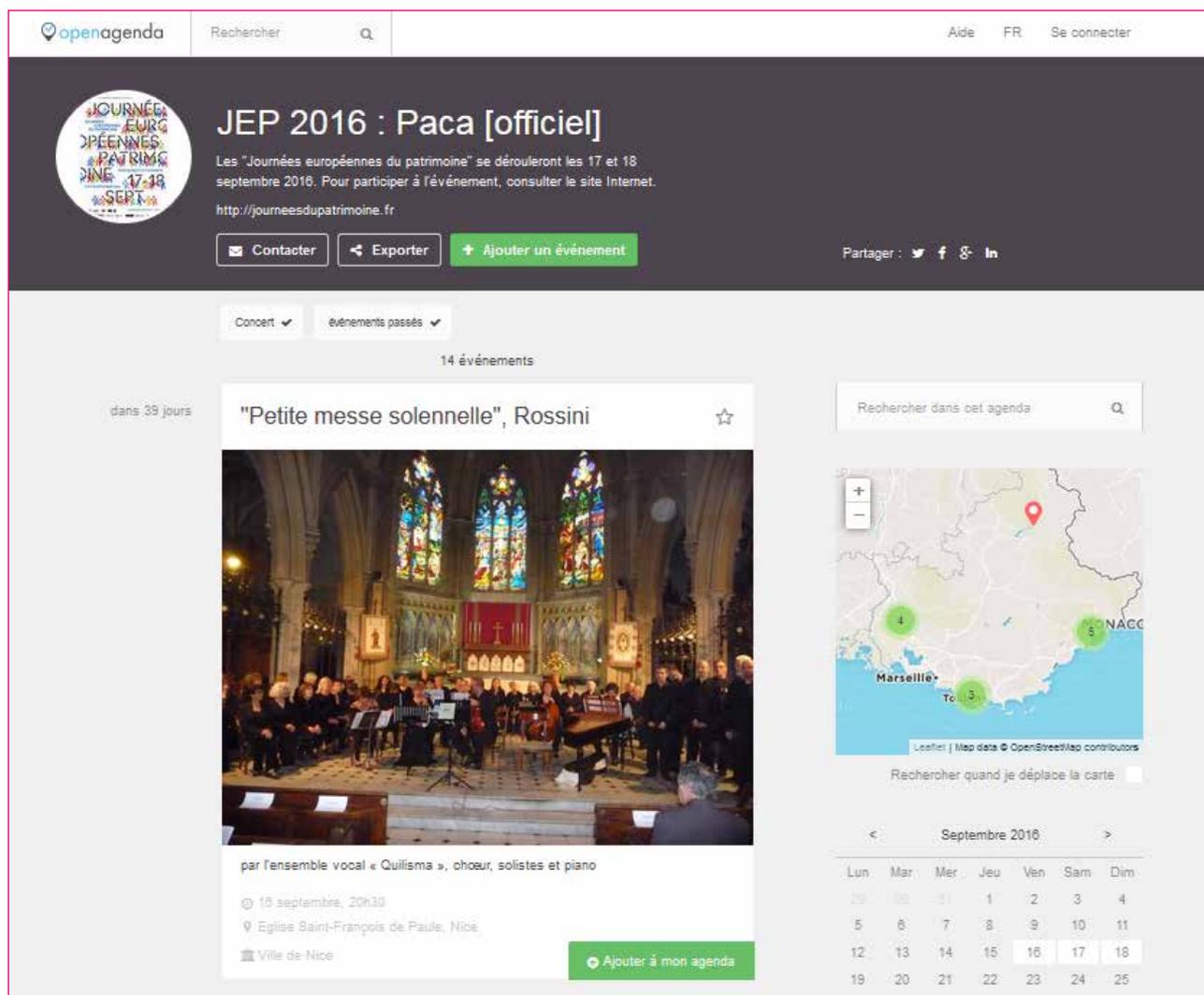


Figure 14 : Interface d'Open Agenda montrant l'ensemble des événements de type Concert dans la région PACA à l'occasion des Journées Européennes du Patrimoine 2016, <https://openagenda.com/jep-2016-paca?oaq%5Bpassed%5D=1&oaq%5Btags%5D%5B0%5D=concert>

Sowprog⁶² est également une plate-forme participative permettant à des acteurs de communiquer rapidement leur programmation culturelle sur de nombreux canaux. La saisie de données événementielles peut se faire à l'aide de simples formulaires, parfois intégrés au système de gestion de contenus tel que Wordpress⁶³. Le programmateur peut alors diffuser simplement ces informations sur plus de 35 agendas culturels, sites web ou applications mobiles partenaires. Les données sont aussi disponibles librement par des ré-utilisateurs. A ce jour, Sowprog revendique avoir près de 30 000 événements enregistrés dans sa base, près de 80% d'entre eux se situant en région parisienne, et il compte parmi ses clients la Philharmonie de Paris.

AgendaCulturel⁶⁴ permet à des professionnels de gérer une structure (un lieu culturel ou un artiste) et de publier des événements à venir. C'est également une interface web permettant de trouver des événements culturels à venir dans une ville ou un département français, se décomposant ainsi dans plus de 100 portails départementaux. Ces événements sont regroupés en 7 catégories : concert, théâtre, festival, jeune public, danse, art du spectacle et exposition. Une application mobile iOS a aussi été développée mais celle-ci ne semble plus disponible. AgendaCulturel revendique lister plus de 150 000 événements culturels à travers la France et monétise l'accès à ses données à travers un abonnement mensuel permettant la réutilisation des données pour une ville ou un département via une API⁶⁵. L'interface web affiche également de nombreuses publicités.

62 - <http://www.sowprog.com/>

63 - <http://www.sowprog.com/focus-pratique-nos-plugins-wordpress-widgets-et-api-pour-vos-sites-web/>

64 - <http://www.agendaculturel.fr/>

65 - <http://pro.agendaculturel.fr/pages/produits/api-agenda-culturel.html>

Mapado⁶⁶ a comme ambition de référencer tous les concerts, spectacles, expositions ou autres événements culturels en France. Pour cela, Mapado collecte depuis 2012 des informations diffusées sur le web, notamment sur les sites proposant des activités tels que les salles de théâtre, les musées, les salles de concert ou encore les clubs de sport. Mapado est donc dans une logique de captation des informations disponibles et déjà publiées sur le Web en utilisant des techniques d'apprentissage et de fouille de données. Mapado permet ensuite aux organisateurs d'événements de mieux les promouvoir en les diffusant rapidement sur plus de 70 agendas culturels parmi les plus consultés. La Figure 14 montre par exemple l'ensemble des événements passés et à venir ayant eu lieu à la cathédrale Sainte Réparate à Nice. On notera que l'interface de recherche affiche souvent des doublons (par exemple, le concert philharmonique « Hors les Murs ») et que les événements prévus à l'occasion des Journées Européennes du Patrimoine au mois de septembre prochain ne sont pas connus. A moyen terme, Mapado envisage d'ouvrir un service de billetterie et d'offrir des services aux animateurs de communauté pour leur permettre de suivre l'attention portée aux événements annoncés sur les réseaux sociaux.

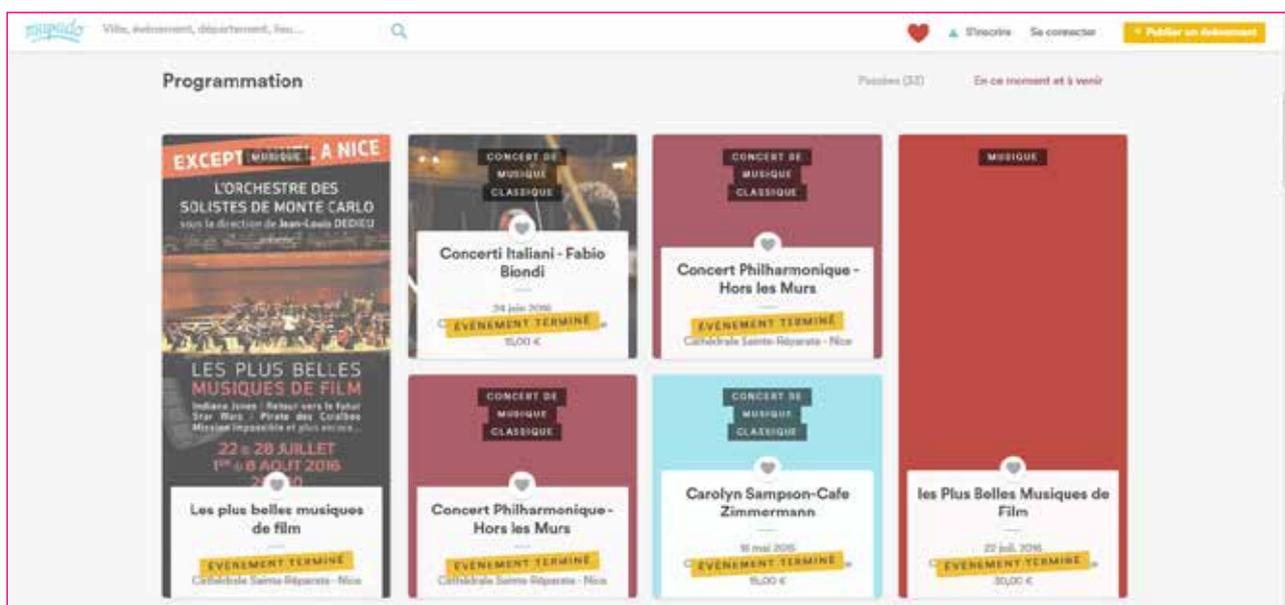


Figure 15 : Interface de Mapado montrant l'ensemble des événements passés qui se sont déroulés à la cathédrale Sainte Réparate à Nice, <https://www.mapado.com/nice/cathedrale-sainte-reparate?futureOrPast=past>

Evensi⁶⁷ est une application web et mobile qui permet de trouver des événements passés et à venir en fonction de ses centres d'intérêts. L'application se nourrit de tous les événements qui sont publiés ouvertement sur les réseaux sociaux, principalement Facebook, et les diffuse à son tour, sous forme de flux. La couverture est mondiale et Evensi revendique une base de 48 millions d'événements. Les événements sont classés en 7 grandes catégories, elles-mêmes sous-divisées en thèmes plus fins. La Figure 15 montre par exemple les événements prévus pour la journée du 16 septembre 2016 à Nice et dans ses environs. On remarque, par exemple, que les événements prévus à l'occasion des Journées Européennes du Patrimoine ne sont pas proposés, ceux-ci n'ayant pas encore fait l'objet d'une diffusion publique sur Facebook.

66 - <https://www.mapado.com/>

67 - <https://www.evensi.com/>

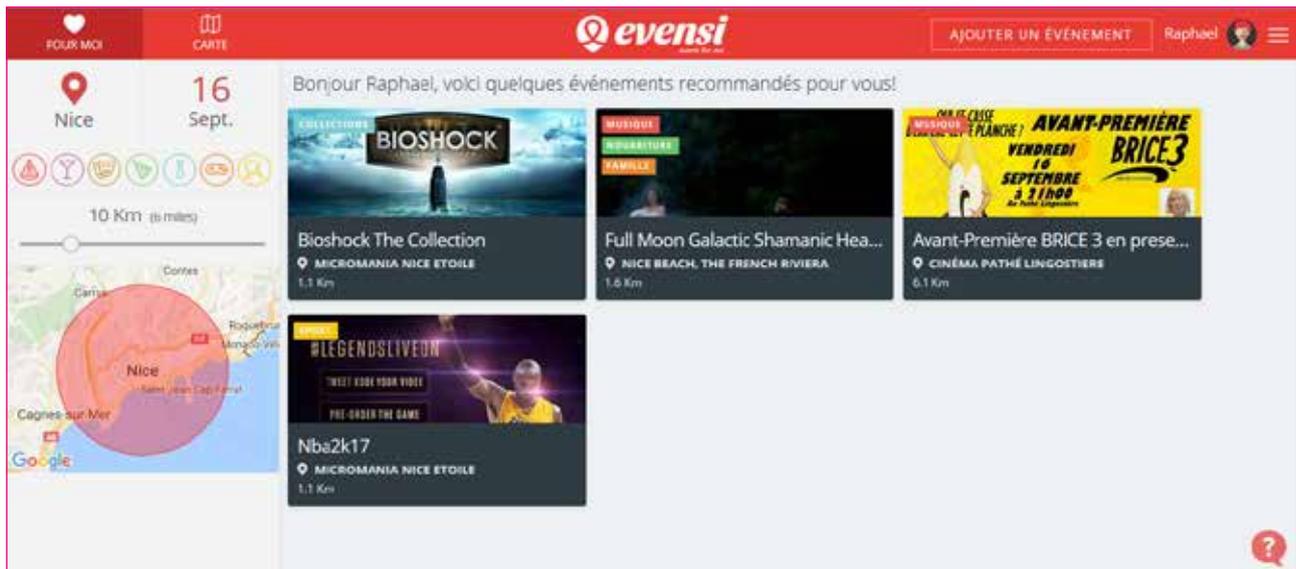


Figure 16 : Interface de Evensi montrant les événements prévus le vendredi 16 septembre 2016 dans un rayon de 10km autour du centre de Nice

Toot Sweet⁶⁸ est une application mobile qui se définit elle-même comme le « Tinder » des sorties. L'application propose des événements proches de chez soi, dans l'instant, selon 5 grandes catégories : soirées, concerts, spectacles, cinés et expos. Il n'est volontairement pas possible de planifier et de découvrir des événements se déroulant dans le futur. L'application propose d'acheter des billets et prélève ainsi une commission. Les données proviennent soit de partenaires qui disposent d'un espace professionnel pour entrer des informations, soit d'acteurs dans la billetterie (France Billet, FNAC, Digitick, etc.). Les données couvrent essentiellement Paris et ses alentours. Si les événements gratuits sont également visibles sur l'application, c'est essentiellement pour des raisons de complétude de l'information. L'application enregistre les j'aime effectués par les utilisateurs sur chacune des sorties qui leurs sont proposées afin d'affiner leur profil et d'améliorer la recommandation. Une application web venant compléter l'application mobile devrait voir le jour à la rentrée.

EventMedia⁶⁹ est un projet de recherche conduit par EURECOM. Il offre une base de connaissances d'événements enrichis sémantiquement et de médias qui leurs sont associés, automatiquement extraits du web (Khrouf, Milicic, & Troncy, 2014). Un moteur de recherche exploratoire permet aux utilisateurs de découvrir des événements à venir, ou de revivre des événements passés à travers les médias (photos et vidéos) qui ont été partagés sur les réseaux sociaux. Une fois qu'un événement est sélectionné, les médias sont présentés pour transmettre l'expérience utilisateur, ainsi que d'autres informations comme la description de l'événement, les artistes ou les utilisateurs qui ont participé (Figure 16). Ce projet a gagné le 1^{er} prix du *Semantic Web Challenge* organisé par la communauté internationale du Web Sémantique en 2012⁷⁰.

68 - <http://tootsweet-app.com/>

69 - <http://eventmedia.eurecom.fr>

70 - <http://challenge.semanticweb.org/2012/>

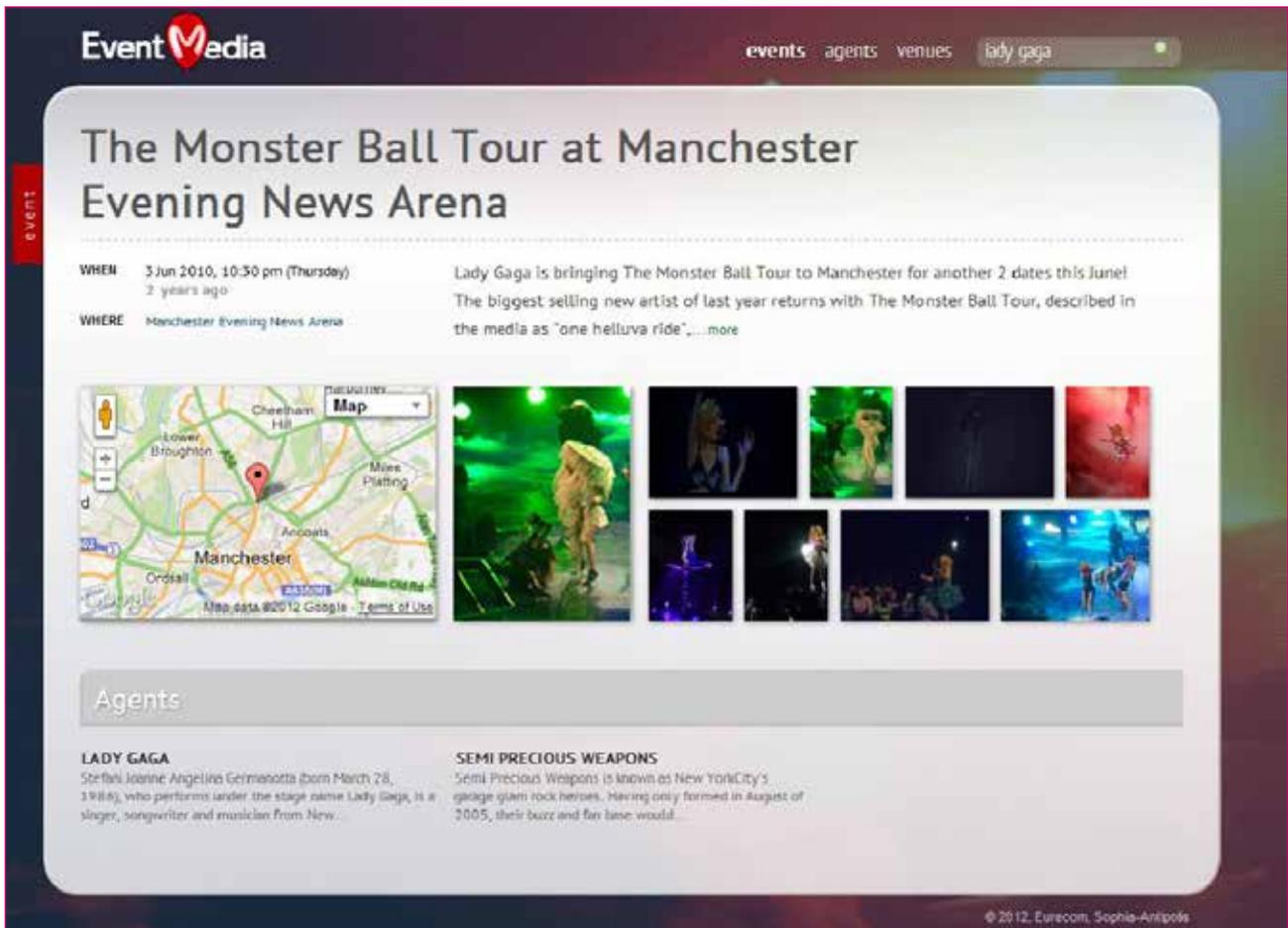


Figure 17 : Interface EventMedia montrant le concert de Lady Gaga à l'Evening News Arena de Manchester dans le cadre de la tournée Monster Ball, le 3 juin 2010

3cixty⁷¹ est un projet d'innovation européen financé par l'EIT Digital qui développe un moteur de recherche exploratoire pour les villes intelligentes. En particulier, il fournit des technologies pour construire des bases de connaissances exhaustives pour des territoires donnés, mêlant informations sur tous les lieux d'intérêts, les commerces, la programmation événementielle et culturelle ainsi que l'offre de transport public disponible. Une interface de recherche et d'exploration permet à des utilisateurs de planifier leurs visites dans la ville en épinglant les éléments qui composeront leurs programmes (Figure 17). Une application mobile sert enfin de compagnon pour aider les utilisateurs à se repérer dans la ville et pour leur rappeler les éléments précédemment vus et épinglés. Milan⁷² fut la première ville couverte par 3cixty à l'occasion de l'exposition universelle qui s'est tenue en 2015. Aujourd'hui, 3cixty fournit une base de connaissances complètes pour les villes de Nice⁷³ et de Londres⁷⁴. Ce projet a gagné le 1^{er} prix du *Semantic Web Challenge* organisé par la communauté internationale du Web Sémantique en 2015⁷⁵.

71 - <https://www.3cixty.com/>

72 - <https://milan.3cixty.com/>

73 - <https://milan.3cixty.com/>

74 - <https://london.3cixty.com/>

75 - <http://challenge.semanticweb.org/2015/>

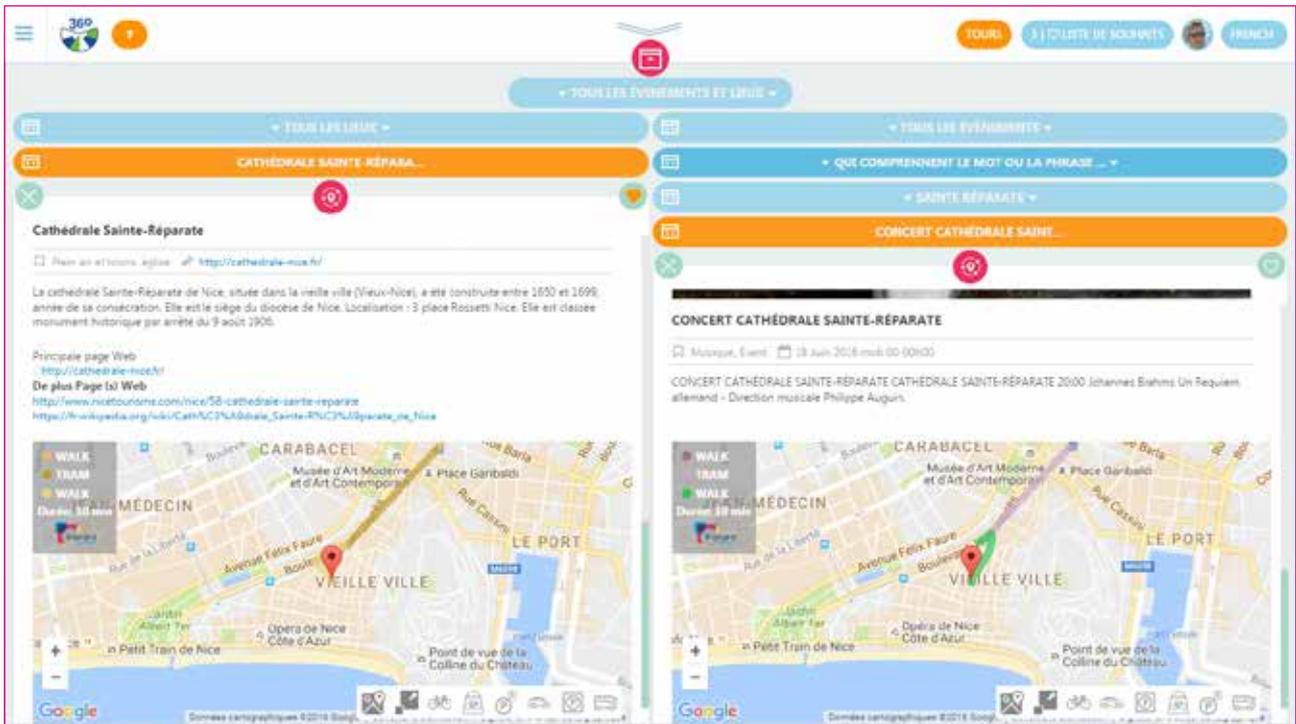


Figure 18 : Interface 3cityxty permettant d'explorer les lieux d'intérêts et la programmation événementielle culturelle pour la ville de Nice. Ici, la cathédrale Sainte-Réparate est sélectionnée, ainsi qu'un événement passé ayant eu lieu le 18 juin dernier

ODE Culture⁷⁶ était un projet lancé par l'association liberTIC, en partenariat avec la société de développement web Makina Corpus⁷⁷, qui avait pour ambition de développer un logiciel libre permettant à une institution de gérer un agenda culturel et en facilitant la réutilisation des données à travers une API standardisée. Le site web de démonstration⁷⁸ n'est plus disponible mais le code du logiciel libre est sur github⁷⁹.

Calendev⁸⁰ se voulait être une application de recommandation d'événements ou de sorties proches de chez soi. L'approche de cette startup toulousaine, lauréate du 4^{ème} concours Dataconnexions⁸¹, se situe donc entre l'annuaire exhaustif listant toute sorte d'événements et le moteur de recherche. L'utilisateur s'abonne à des catégories assez fines d'événements ou à des lieux et peut alors suivre sous forme de flux continu d'information l'agenda culturel pour ensuite partager ses coups de coeur. Lancée sous forme d'application mobile, ou sur le web, Calendev était donc une véritable application de type SoLoMo⁸² par son approche marketing. Gratuite tant pour les utilisateurs que pour les organisateurs d'événements, elle paraît sur un modèle d'affaire liée à la publicité, qui devait être largement présente dans l'application. Cependant, l'application n'est plus disponible, confirmant la faiblesse du modèle d'affaire. La stratégie à long terme que l'application comptait utiliser pour nourrir sa base d'événements et de sorties à proposer, au-delà des données publiques disponibles sur les portails de données des collectivités, n'est pas claire.

76 - <http://fr.slideshare.net/libertic/lancement-projet-ode-culture>

77 - <http://makina-corporus.com/>

78 - <http://projet-ode.org/>

79 - <https://github.com/LiberTIC/ODEV2>

80 - <http://www.calendev.com/>

81 - <https://www.etalab.gouv.fr/dataconnexions4decouvrezeshuitlaureatsdecetteedition>

82 - L'acronyme SoLoMo est la contraction des termes Social, Local et Mobile, https://fr.wikipedia.org/wiki/Marketing_%C3%A9lectronique

Pour être tout à fait complet, mentionnons également les nombreux annuaires d'événements à portée mondiale, plus ou moins spécialisés, et qui couvrent nécessairement la France, tels que : Eventful⁸³, Eventbrite⁸⁴, Last.fm⁸⁵ ou plus récemment WithEvent⁸⁶. Ces annuaires offrent généralement une API permettant d'accéder aux données même si les conditions d'utilisation interdisent en général le développement d'application commerciale. Ces annuaires listent principalement des événements d'importance susceptibles de rassembler un public large mais ont souvent une couverture locale très pauvre. Là est le paradoxe : la connaissance exhaustive de l'offre culturelle pour un territoire donné est souvent enfermée à des acteurs locaux qui ne disposent pas d'outils pour augmenter la visibilité et la diffusion des événements culturels alors que les utilisateurs ont massivement recours à des moteurs de recherche grands public ou à des applications généralistes qui ne proposeront pas d'événements hyper locaux. Ce travail de fourmi consistant à collecter l'offre culturelle est ainsi souvent laissée à des collectifs⁸⁷, le travail de l'association Jondi à l'occasion de la dernière fête de la musique en est un exemple⁸⁸.

—

83 - <http://eventful.com/>

84 - <https://www.eventbrite.com/>

85 - <http://www.last.fm/>

86 - <http://withevent.com/>

87 - <http://fabrikasons.com/>

88 - <http://jondi.fr/actualites/fete-de-la-musique-ministere-sa-mere/>

AVIS, INTERACTIONS SOCIALES ET RECOMMANDATIONS

Le Web est un espace social où les utilisateurs sont invités à interagir. FOAF (de l'anglais *Friend of a Friend*) est une ontologie RDF permettant de décrire des personnes et les relations qu'elles entretiennent entre elles. Cette ontologie⁸⁹ est un élément central du web sémantique et est souvent utilisée comme référence par plusieurs centaines d'autres vocabulaires.

Les avis et critiques postés par les utilisateurs, attachés à un produit culturel, à un lieu, à un artiste ou à un événement, constituent un premier exemple d'interactions sociales. Historiquement, le micro-format *hReview*⁹⁰ (puis *h-review*⁹¹) s'est imposé pour représenter dans une page web sous forme structurée une critique composée d'un résumé et d'une note sur une échelle normalisée. Le vocabulaire *schema.org* définit également la classe *schema:Review* qui peut inclure un *schema:Rating*.

D'autres types d'interactions sociales sont aussi populaires telles que le fait de donner une réaction à un contenu publié sur le web (*Like, Dislike, Flag, Ignore*) ou encore le fait de répondre à une invitation pour participer à un événement (*Accept, Ignore, Reject*). Le groupe de travail Social Web du W3C définit un vocabulaire⁹² composé de classes et de propriétés listant toutes ces interactions sociales. *Schema.org* définit lui aussi ces interactions possibles en introduisant le concept d'actions qui permettent d'interagir avec des contenus directement depuis d'autres services tels que Gmail et Inbox. Ainsi, les organisateurs d'événements culturels peuvent inclure des métadonnées sémantiques supplémentaires permettant aux utilisateurs d'indiquer leur participation avec la classe *schema:RsvpAction*.

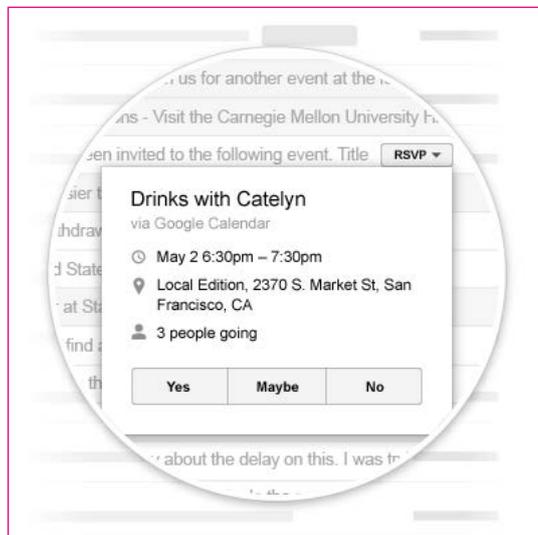


Figure 19 : Exemple d'intégration des actions de *schema.org* dans l'application Gmail où l'utilisateur peut directement indiquer sa présence à un événement depuis un courriel

Les avis et opinions postés par les utilisateurs servent souvent aux moteurs de recommandation qui ont pour objectif de réduire la surcharge d'information et de guider l'utilisateur à prendre une décision qui correspond à ses intérêts. Dans un service qui fournit des milliers d'événements par jour, les options de navigation deviennent insuffisantes et un moteur de recommandation s'avère indispensable pour optimiser l'expérience utilisateur. En particulier, la recommandation d'un événement met en jeu plusieurs facteurs comme le temps, le lieu, la popularité des artistes, un profil utilisateur composé de centres d'intérêts voire la confirmation de la participation d'« amis » à cet événement.

La recommandation basée sur le contenu ou la recommandation thématique s'appuie sur le contenu des objets pour proposer des profils similaires à ceux qui ont été précédemment appréciés par l'utilisateur (Pazzani, Billsus, 2007). Le système compare le profil d'un objet (composé de descripteurs) avec d'autres profils intéressants afin de prédire l'opinion de l'utilisateur sur cet objet. Cette comparaison consiste à calculer la similarité entre les profils des objets, ce qui peut être mesuré par le biais de plusieurs métriques comme la similarité de Pearson ou la similarité Cosinus. Pour représenter le profil d'un objet, la méthode la plus commune est la représentation des métadonnées en utilisant TF-IDF (Term Frequency-Inverse Document Frequency) (Généreux, Santini, 2007) qui permet d'évaluer un mot-clé par sa fréquence dans un document et par sa présence dans tous les autres documents du corpus.

89 - <http://xmlns.com/foaf/0.1/>

90 - <http://microformats.org/wiki/hreview>

91 - <http://microformats.org/wiki/h-review>

92 - <https://www.w3.org/TR/activitystreams-vocabulary/>

Une telle représentation nécessite des techniques d'extraction des caractéristiques pour transformer une description non structurée en une forme structurée soulignant les métadonnées. Cette extraction devient extrêmement simple avec les technologies de web sémantique grâce à la structuration des données dans des ontologies. En outre, le modèle sémantique permet d'enrichir la description d'un objet avec des informations supplémentaires en exploitant la richesse du web des données.

Cependant, cette approche ne prend pas en compte la dimension sociale qui répond à la question « quels sont mes amis qui vont assister à un événement donné ? ». Pour apporter une réponse, la recommandation basée sur le filtrage collaboratif permet de mettre en avant la dimension sociale. Cette technique a été popularisée par Amazon avec sa fonctionnalité « les gens qui ont acheté X ont aussi acheté Y ». La méthode consiste à chercher des utilisateurs qui ont le même comportement avec l'utilisateur dont on souhaite faire des recommandations et d'utiliser les notes et avis, ou les interactions sociales de ces utilisateurs similaires pour calculer une liste de recommandations.

Aujourd'hui, les moteurs de recommandation sont souvent hybrides et utilisent à la fois des techniques calculant la similarité entre les objets et des techniques de type filtrage collaboratif.

—

CONCLUSION

Nous avons présenté dans ce document en quoi consistait les technologies du web sémantique ou orientées 3.0 en mettant l'accent sur la disponibilité croissante de bases de connaissances, elles-mêmes utilisées par les moteurs de recherche généralistes offrant ainsi des visualisations sous forme de cartes de connaissances directement dans les pages web de résultats. Nous avons longuement décrit le vocabulaire schema.org, qui s'est imposé ces dernières années comme un standard pour représenter des données structurées sur le web. Il bénéficie d'une communauté active qui garantit son développement, la dernière version 3.1⁹³ ayant été publiée le 9 août 2016. Nous avons également mis en avant le fait que le support complet de ce vocabulaire complexe par les outils informatiques était encore insuffisant (par exemple, l'outil de test de Google ne supporte que la version 2.2 de schema.org) et nous avons proposé un outil open source pour pré-visualiser des métadonnées riches décrivant des œuvres et des événements culturels. Un meetup⁹⁴ consacré à schema.org a eu lieu à Paris le 6 octobre 2016 : il a permis de faire le point sur les développements en cours, sur les outils qui produisent et qui consomment ces métadonnées et l'impact que cela a dans le référencement sur le web⁹⁵.

L'application des technologies du web sémantique à l'offre du secteur culturel renforcera sans aucun doute la visibilité de ses différents acteurs (artistes, créateurs, producteurs) sur le web et démultipliera les formes de valorisation de la création. Du côté du public, d'autres formes d'interactions seront possibles avec, notamment, le suivi de l'activité des artistes. Pour les centres de ressources culturelles, ces technologies permettent de réduire les saisies manuelles et de récupérer des métadonnées descriptives. Enfin, elles augmenteront la présence de ces acteurs dans les résultats des moteurs de recherche généralistes car ils seront inclus dans leurs graphes de connaissances.

Les technologies sémantiques sont particulièrement adaptées à l'internationalisation des données notamment à des fins de multilinguisme. Cet enjeu concerne au premier plan le tourisme culturel international. Pour des raisons évidentes de coût de traduction, les agendas des établissements sont rarement multilingues. L'internationalisation des données de programmation augmenterait la visibilité de l'offre auprès du public étranger tout en favorisant des dynamiques interculturelles. Nous avons déjà mentionné l'expérience de JocondeLab réalisée en partenariat avec le ministère de la Culture et de la Communication en 2014 qui a permis de donner un accès en quatorze langues (dont le chinois et le japonais) à quelque 350 000 notices d'œuvres extraites de la base Joconde.

Nous concluons ce document par un certain nombre de préconisations qui peuvent préfigurer d'autant d'actions à mener à court et moyen termes :

Préconisation 1 : développer un référentiel standard et multilingue pour catégoriser les événements. Il s'agit souvent du premier⁹⁶ filtre que les applications utilisent pour présenter ou recommander des agendas culturels. Un tel vocabulaire devrait être lisible par les machines, ouvert, et refléter les multiples dimensions permettant de catégoriser un événement, entre genre, style ou public visé.

Préconisation 2 : encourager les institutions culturelles à produire et publier des métadonnées Twitter, Facebook et schema.org pour faciliter le partage des informations sur les réseaux sociaux.

Préconisation 3 : adopter le modèle de données schema.org, compatible avec les technologies du web sémantique, plutôt que d'avoir recours à un modèle de données ad-hoc et propriétaire.

93 - <http://schema.org/docs/releases.html>

94 - <https://www.meetup.com/fr-FR/paris-web-of-data/events/232635663/>

95 - <https://www.youtube.com/watch?v=bOLBGO-Xds>

96 - [La géolocalisation étant souvent automatique, on ne le considère pas ici comme un filtre](#)

Préconisation 4 : automatiser la publication des agendas culturels, sous forme de jeux de données ayant une licence libre permettant la réutilisation, mis à jour régulièrement, via des fichiers et/ou des APIs et sortir d'une logique d'archivage de ces données qui perdent de leur valeur une fois que l'événement s'est déroulé.

Préconisation 5 : utiliser les technologies sémantiques pour analyser et dédoublonner les métadonnées décrivant les événements culturels et les objets associés, en adoptant par exemple la vision portée par le projet 3cixty.

Ces préconisations pourront être mises en oeuvre à l'occasion de la refonte de l'agenda culture qui verra le jour en 2017⁹⁷. De plus, conformément aux conclusions de la note de cadrage issue des réflexions du groupe de travail n°7 préfigurant à ce rapport, il semble opportun de réaliser un prototype d'outil d'édition de données de programmation culturelle fondées sur les technologies du web sémantique. Les établissements volontaires du Parc de la Villette constitueraient un pôle interinstitutionnel idéal pour expérimenter un tel prototype et la Philharmonie de Paris - de par son expérience sur le développement de la plateforme MIMO ainsi qu'au sein du projet DOREMUS - serait la mieux à même pour assurer la maîtrise d'ouvrage de ce projet pilote.

REMERCIEMENTS

L'auteur tient particulièrement à remercier Marie-Véronique Leroi et Bertrand Sajus pour leurs nombreux commentaires sur des versions précédentes de ce document, tous les membres du groupe de travail GT7 mis en place par le Ministère de la Culture et de la Communication (Rodolphe Bailly, Yannick Vernet, Sandrine Heiser, Enora Prioul et Anne Texeira), Romain Lange à l'initiative d'Open Agenda, Aly Micki directeur technique du produit Sowprog, Hatem Chelbi co-fondateur de la startup TootSweet, Lucile Colombain responsable du projet Open Data Events, Fabien Gandon, directeur de recherche à l'Inria, Jean Delahousse, ainsi que Houda Khrouf et Pasquale Lisena qui ont travaillé sur les outils mentionnés dans ce rapport.

97 - <http://www.culturecommunication.gouv.fr/L-agenda-culture.fr-evolue>

GLOSSAIRE

API : *Application Programming Interface* (ou interface de programmation applicative) est un ensemble normalisé de classes, de méthodes ou de fonctions qui sert de façade par laquelle un logiciel offre des services à d'autres logiciels. Les APIs Web représentent un type d'API pour lequel une entreprise permet à des applications tierces d'utiliser, par exemple, ces données.

CSV : *Comma-Separated Values* est un format informatique ouvert permettant de représenter des données tabulaires sous formes de valeurs séparées par des virgules. Un fichier CSV est un fichier texte par opposition aux formats dits binaires. Chaque ligne du texte correspond à une ligne du tableau et les virgules correspondent aux séparations entre les colonnes.

FRBRoo : *Functional Requirements for Bibliographic Records Object Oriented* est une ontologie formelle qui a comme ambition de représenter la sémantique des informations bibliographiques dans le but de faciliter l'intégration et l'échange de données entre les bibliothèques et les musées.

JSON-LD : *JavaScript Object Notation for Linked Data* est une méthode permettant de représenter des données RDF dans le format JSON largement utilisé par les développeurs Web. C'est une recommandation du W3C depuis 2014, <https://www.w3.org/TR/json-ld/>.

Ontologie : en informatique ou en science de l'information, une ontologie est un modèle de données, un ensemble structuré de termes, de concepts et de relations représentant le sens des éléments d'un domaine de connaissances. Schema.org est un exemple d'ontologie adopté par les principaux moteurs de recherche du Web.

Métadonnées : ce sont littéralement des données servant à définir ou à décrire d'autres données. Héritage des établissements tels que les bibliothèques, médiathèques ou archives qui ont une longue pratique dans la codification du signalement ou des contenus des documents (les notices bibliographiques), le terme métadonnée est apparu dans le cadre de la description de ressources sur le Web et s'est ensuite généralisé.

RDF : *Resource Description Framework* est un modèle de graphe permettant de représenter des données et des métadonnées associées à des ressources Web. RDF est la pierre angulaire du Web Sémantique. C'est une recommandation du W3C depuis 1999, <http://www.w3.org/TR/rdf11-primer/>.

RDFa : *Resource Description Framework in Attributes* définit une syntaxe permettant d'ajouter des données structurées en RDF directement dans les éléments HTML d'une page Web. Le code RDFa est invisible pour l'internaute et n'affecte pas ce qui est affiché dans un navigateur, mais il est utile aux robots d'indexation ou autres applications qui peuvent utiliser ces données. C'est une recommandation du W3C depuis 2008, <https://www.w3.org/TR/rdfa-primer/>

Schema.org : vocabulaire constitué de classes et de propriétés permettant de décrire sémantiquement le contenu de pages Web. Initialement porté par les principaux moteurs de recherche du Web, le vocabulaire évolue aujourd'hui de manière ouverte grâce aux contributions d'une communauté de développeurs. De nombreux outils permettent de générer⁹⁸ ou de valider⁹⁹ des métadonnées schema.org.

SKOS : *Simple Knowledge Organization System* s'appuie sur le modèle de données RDF pour représenter des thésaurus, des classifications ou d'autres types de vocabulaires contrôlés. Les vocabulaires SKOS sont conçus pour être intégrés dans des environnements sémantiques utilisant par exemple des ontologies OWL. C'est une recommandation du W3C depuis 2009, <https://www.w3.org/TR/skos-primer/>.

XML : *eXtended Markup Language* est une famille de langage informatique de balisage générique dont l'objectif est de faciliter l'échange automatisé de contenus entre systèmes d'informations.

98 - Schema.org generator : <http://schema.pythonanywhere.com/> ou <https://www.schemaapp.com/schema-org-json-ld-markup-generator/>

99 - Structured Data Linter: <http://linter.structured-data.org/>; Google Structured Data Testing Tool: <https://search.google.com/structured-data/testing-tool/>; Yandex Structured Data Validator: <https://webmaster.yandex.com/microtest.xml>

RÉFÉRENCES

- Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The Semantic Web. *Scientific American*, 29-37.
- Doerr, M. (2003). The CIDOC Conceptual Reference Module: An Ontological Approach. *AI Magazine*, 75-92.
- Gangemi, A., & Mika, P. (2003). Understanding the Semantic Web through Descriptions and Situations. 2nd International Conference on Ontologies, *Databases and Applications of SEMantics (ODBASE)*, (pp. 689-706). Catania, Italy.
- Guha, R. V., Brickley, D., & Macbeth, S. (2016). Schema.org: Evolution of Structured Data on the Web. *Communications of the ACM*, 44-51.
- Khrouf, H., & Troncy, R. (2014). De la modélisation sémantique des événements vers l'enrichissement et la recommandation. *Revue d'Intelligence Artificielle (RIA)*, 321-347.
- Khrouf, H., & Troncy, R. (2011). Réconcilier les événements dans le web de données. *22nd Journées Francophones d'Ingénierie des Connaissances (IC)*. Chambéry, France.
- Khrouf, H., Milicic, V., & Troncy, R. (2014). Mining Events Connections on the Social Web: Real-Time Instance Matching and Data Analysis in EventMedia. *Journal of Web Semantics (JWS)*, 3-10.
- Lagoze, C., & Hunter, J. (2001). The ABC Ontology and Model. *Journal of Digital Information (JoDI)*.
- Norvig, P. (2016). The Semantic Web and the Semantics of the Web: Where Does Meaning Come From? *25th World Wide Web Conference*, (p. 1). Montreal, Canada.
- Raimond, Y., Abdallah, S., Sandler, M., & Giasson, F. (2007). The Music Ontology. *8th International Conference on Music Information Retrieval (ISMIR)*. Vienna, Austria.
- Scherp, A., Franz, T., Saathoff, C., & Staab, S. (2009). F - A Model of Events based on the Foundational Ontology DOLCE+ Ultra Light. *5th International Conference on Knowledge Capture (K-CAP)*. Redondo Beach, California, USA.
- Troncy, R., Shaw, R., & Hardman, L. (2010). LODÉ: une ontologie pour représenter des événements dans le web de données. *21th Journées Francophones d'Ingénierie des Connaissances (IC)*, (pp. 69-80). Nîmes, France.
- Van Hage, W., Malaisé, V., de Vries, G., Schreiber, G., & van Someren, M. (2009). Combining Ship Trajectories and Semantics with the Simple Event Model (SEM). *1st ACM International Workshop on Events in Multimedia (EiMM)*. Beijing, China.
-

ANNEXE A :
DESCRIPTION DE L'ÉVÉNEMENT « LIBERTÉ ! » À L'AIDE DE SCHEMA.ORG

```

[[
  «@context»: «http://schema.org/»,
  «@type»: «MusicEvent»,
  «description»: «Concert par l'ensemble vocal \»Ristretto\» et le \»BBC Symphony Chorus\»\n\n« Liberté »\n«
  Figure Humaine », de Francis Poulenc et œuvres de Hector Berlioz, Claude Debussy, Claudio Monteverdi,
  Hugo Wolf par l'ensemble vocal « Ristretto » et chanteurs du \»BBC Symphony Chorus\»\nDirection Errol
  Girdlestone «,
  «image»: {
    «@type»: «ImageObject»,
    «contentUrl»: «https://api.getwemap.com/images/pics-picpoints/b18b384a6da16d80dad46e90.04427439.
    jpg?width=500&always_fit=false»,
    «creator»: «Ristretto»
  },
  «location»: {
    «@type»: «Place»,
    «name»: «Cathédrale Sainte-Réparate»,
    «address»: {
      «@type»: «PostalAddress»,
      «addressCountry»: «France»,
      «addressLocality»: « Nice»,
      «postalCode»: «06000»,
      «streetAddress»: «Place Rossetti»
    }
  },
  «performer»: [{
    «@type»: «PerformanceRole»,
    «roleName»: «Director»,
    «performer»: {
      «@type»: «Person»,
      «name»: «Errol Girdlestone»,
      «image»: «http://marc.syl20.free.fr/images/errolGirdlestone.jpg»µ
    }
  }, {
    «@type»: «PerformanceRole»,
    «roleName»: «Chorus»,
    «performer»: {
      «@type»: «MusicGroup»,
      «name»: «Ensemble vocal \»Ristretto\»»,
      «image»: «https://scontent-bru2-1.xx.fbcdn.net/v/t1.0-
      9/11539781_799484743506088_6237680323271273838_n.png?oh=6c7d40e65935e23172dffbc-
      376b364ac&oe=58243D0E»
    }
  }, {
    «@type»: «PerformanceRole»,
    «roleName»: «Chorus»,
    «performer»: {
      «@type»: «MusicGroup»,
      «name»: «BBC Symphony Chorus»,
      «image»: «https://encrypted-tbn1.gstatic.com/images?q=tbn:ANd9GcSXmvc7Nf4CRFTThkGezHPbfQ2wj-
      GuqIRY0et5dN8a8NQVtYXDgZRQ»
    }
  }
  ]],
  «name»: «Liberté !»,
  «startDate»: «2016-09-17T20:00»,
  «endDate»: «2016-09-17T22:00»,
  «doorTime»: «20:00»,
  «workPerformed»: {
    «@id»: «bnode0»
  }

```

```

},
«url»: «http://journeesdupatrimoine.culturecommunication.gouv.fr/Programme?logid=1735682»,
«offers»: [{
  «@type»: «Offer»,
  «name»: «Entrée libre»,
  «price»: 0,
  «priceCurrency»: «EUR»,
  «url»: «https://citm-philharmoniedeparis.shop.secutix.com/api/1/redirect/product/perfor-
mance?id=618710145&lang=en»
}]
}, {
  «@context»: «http://schema.org/»,
  «@type»: «Event»,
  «description»: «Des vestiges les plus anciens au patrimoine du XXIe siècle : parcourez les premiers monu-
ments figurant dans la programmation qui s'enrichira au fil de l'été avec de nombreuses ouvertures et des
animations inédites.»,
  «name»: «Journées Européennes du Patrimoine»,
  «startDate»: «2016-09-17»,
  «endDate»: «2016-09-18»,
  «image»: «https://pbs.twimg.com/media/CoNfX6ZXgAAN8si.jpg»,
  «url»: «http://journeesdupatrimoine.culturecommunication.gouv.fr/»
}, {
  «@context»: «http://schema.org/»,
  «@type»: «MusicComposition»,
  «@id»: «:bnode0»,
  «image»: «https://upload.wikimedia.org/wikipedia/commons/thumb/a/a7/Eug%C3%A8ne_Delacroix_-_La_
libert%C3%A9_guidant_le_peuple.jpg/500px-Eug%C3%A8ne_Delacroix_-_La_libert%C3%A9_guidant_le_
peuple.jpg»,
  «name»: «Figure Humaine»,
  «description»: «Figure humaine de Francis Poulenc est une cantate pour double chœur mixte composée en
1943 sur des textes de Paul Éluard, créée à Londres en anglais par la BBC en 1945, en français en 1946 à
Bruxelles et en France le 22 mai 1947 à Paris.»,
  «dateCreated»: «1943»,
  «agent»: {
    «@type»: «Role»,
    «roleName»: «composer»,
    «agent»: {
      «@id»: «:poulenc»
    }
  }
},
«timeRequired»: «PT18M»,
«musicCompositionForm»: «Cantate»
}, {
  «@context»: «http://schema.org/»,
  «@id»: «:poulenc»,
  «@type»: «Person»,
  «familyName»: «Poulenc»,
  «givenName»: «Francis»,
  «image»: «https://upload.wikimedia.org/wikipedia/en/b/b8/Poulenc-1922.jpg»,
  «name»: «Francis Poulenc»,
  «birthDate»: «1899-01-07»,
  «deathDate»: «1963-01-30»
}]

```



APPLICATION DES TECHNOLOGIES 3.0 À LA DESCRIPTION DES ÉVÉNEMENTS CULTURELS



Publié en décembre 2016

Ce document est mis à disposition sous licence CC BY-SA 3.0 FR
(<https://creativecommons.org/licenses/by-sa/3.0/fr/>)

