Une ressource publiée sur tmnlab.com

Cap Data – France 2030 : retour sur un cas d'usage Découvrabilité et territoire

Posted on 19 octobre 2026 by Roxane Tchernia

Pionnier en France, initié par la Réunion des Opéras de France, piloté par Eudes Peyre jusqu'à la fin 2024, le projet CapData Opéra* – France 2030, porte sur l'industrialisation d'une solution de structuration et de diffusion des données culturelles capable de répondre aux enjeux de développement des publics, de soutien à la création artistique et d'accès à la culture.

A l'origine, ce projet réunissait 8 partenaires, dont 6 maisons d'opéra : Opéra national de Bordeaux, Théâtre du Châtelet, Opéra de Limoges, Opéra Comique, Opéra national du Capitole de Toulouse, Opéra de Rennes et le réseau du TMNlab / laboratoire Théâtres & Médiations Numériques. En 2025, le TMNlab, en collaboration avec la Réunion des Opéras de France, reprend le pilotage du projet pour mener un audit. Objectif ? Identifier de possibles cas d'usage à expérimenter au cours de la dernière année du projet.

En juillet 2025, le TMNlab publie un état des lieux et des recommandations. Puis engage à la rentrée 2025, en convergence d'une démarche coportée avec le Ministère de la Culture, une expérimentation autour de la **découvrabilité territoriale sur la Métropole de Lyon**.

Retour sur l'expérimentation, ses perspectives et ses limites.

Cap Data est soutenu par l'État dans le cadre du dispositif « Expérience augmentée du spectacle vivant » de France 2030 opéré par la Caisse des Dépôts.

Au sommaire :

- Constats de départ
- Qu'est-ce que CapData ?
- L'expérimentation Découvrabilité et territoire
- Les premiers apprentissages
- Ressources et pistes pour la suite

AGENDA
Une après-midi dédiée à la découvrabilité du spectacle vivant et au projet Cap Data – France 2030 se tiendra le 30 novembre 2026 à Paris lors des *Rencontres TMNlab* | *Médiation, attention, découvrabilité : nouvelles responsabilités culturelles à l'ère numérique*
Programme et inscription

Constats de départ

Les structures culturelles ont souvent à renseigner, manuellement, les informations de leurs spectacles sur de multiples systèmes : leurs outils de billetterie, leurs sites internet et toutes les plateformes tierces de diffusion de ces informations, à l'image des bases de tourisme. Des outils avec des champs, des formats et des référentiels non uniformes, ce qui alourdit et complique les tâches de saisie.

À l'inverse, l'offre de programmation de spectacle vivant semble parfois insuffisamment représentée dans les bases de tourisme, bien que les offices de tourisme contribuent souvent à enrichir manuellement leurs propres sites avec les données de spectacles d'organisations partenaires.

Le 2 décembre, lors de l'atelier de lancement de la démarche autour d'un cadre d'interopérabilité pour le spectacle vivant, des participants (parmi lesquels l'Opéra de Lyon, l'Office de tourisme de Lyon, Apidae) ont mis en avant la nécessité d'automatiser la remontée des données de spectacles depuis le site internet d'un lieu vers celui de l'office du tourisme pour mieux les valoriser sans augmenter les actions manuelles de saisies, chronophages et sources d'erreur.

Un groupe de travail s'est ensuite constitué autour de cette idée, renforcé par la présence de l'agence Rezo Zero, qui avait participé au projet CapData et d'ADN Tourisme, la Fédération nationale des organismes institutionnels de tourisme.

DATAtourisme par ADN Tourisme

Lancée en 2018, DATAtourisme est une plateforme d'agrégation à un niveau national de la donnée touristique locale de l'ensemble du territoire, afin de favoriser sa diffusion pour de multiples types de réutilisations.

À travers une ontologie alignée avec Schema.org, la plateforme portée par la Fédération ADN Tourisme réunit de façon structurée les données des offices de tourisme, agences départementales et comités régionaux, en partie par le biais de systèmes d'information touristique (SIT), comme Apidae.

Ces données sont ensuite diffusées en *open data* avec une licence ouverte (le créateur de la donnée doit être cité).

Qu'est-ce que CapData ?

> *J'écris depuis un moment sur la façon dont la visibilité numérique est mal comprise — surtout dans les secteurs des arts, de la culture et du public. On investit dans*

> *des campagnes, des stratégies de contenu, mais on ignore l'élément dont tout dépend : la structure. La découvrabilité est un acte d'agentivité. Les métadonnées sont une infrastructure. Mais rien de tout cela ne compte si les gens ne peuvent pas s'en servir concrètement.*
> Tammy Lee, CEO, Culture Creates | Semantic Technology for the Arts – Québec

CapData est un projet France 2030 initié par la Réunion des Opéras de France et délégué début 2025 au TMNlab qui en était partenaire.

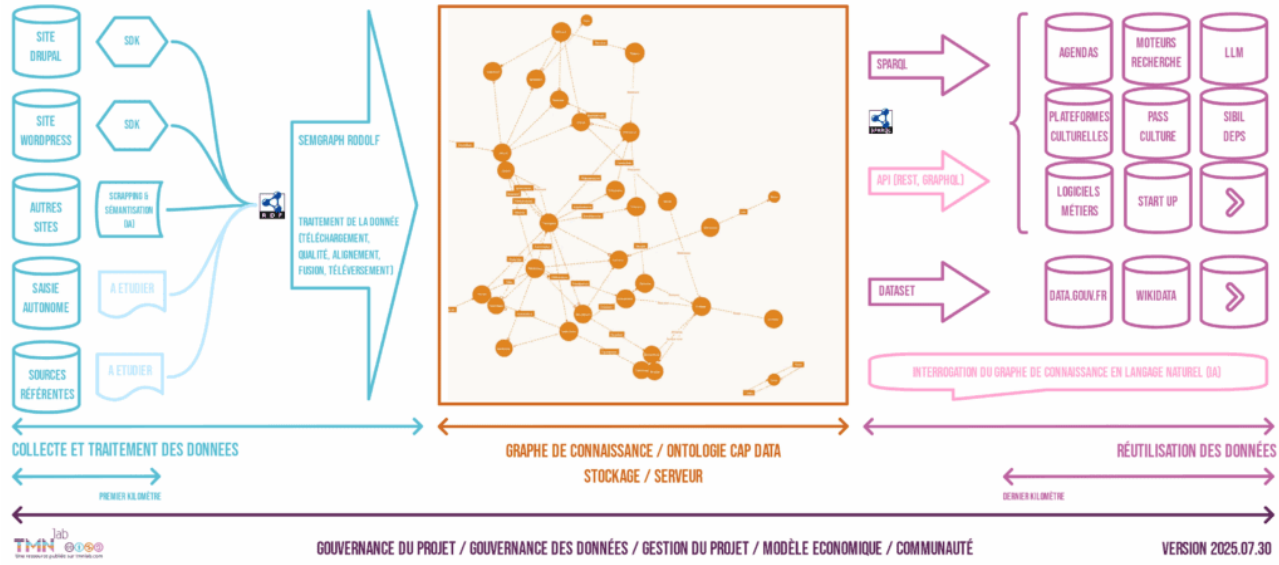
Au départ, il visait à créer un graphe de connaissances autour de l'opéra, s'appuyant sur une ontologie de description spécifique, pour permettre d'agréger les données des maisons d'opéra afin d'ouvrir des possibilités d'exploitation de ces données de qualité. Une preuve de concept (POC) a été réalisée et a permis de produire :

- l'ontologie CapDataCulture
- un graphe de connaissances
- des plugins d'exposition des données dans un format aligné à CapDataCulture (l'un pour Drupal, l'autre pour WordPress)

> *Tout en favorisant la mutualisation, le choix d'une architecture permet une grande agilité, assure le respect de la souveraineté et la pleine responsabilité des maisons, partenaires et services externes participant au projet. Ils conservent une complète autonomie dans le choix d'ouverture, de diffusion, de réutilisation et de l'enrichissement de leurs données.*
>
> E Peyre, F Amarger, N Chauvat

Très vite, il est apparu que ce projet pouvait étendre sa portée aux arts vivants dans leur ensemble. De l'audit mené par le TMNlab à l'été 2025, sont ressortis des points très positifs, comme la puissance descriptive d'un graphe de données, l'intérêt de produire pour la collecte des plugins ouverts pour les principaux CMS ou la richesse des opportunités entrevues. Des pistes d'amélioration et d'approfondissement ont également été révélées :

- la diversification des modes de collecte des données ;
- le développement de cas d'usage d'exploitation de la donnée en sortie ;
- l'ouverture vers une ontologie plus large (pour mieux représenter les arts vivants dans leur globalité) et standardisée (pour un maintien dans le temps et une meilleure appropriation).



Le socle CapData en place, avec ses perspectives d'évolution, et le projet de remontée automatique des données vers les sites d'Office de tourisme se rencontrant, l'agence Logilab à l'origine du graphe de connaissances CapDataCulture, et la solution Ideactiv, qui agrège déjà de nombreuses données culturelles pour les offices de tourisme, ont rejoint le groupe de travail dédié à cette expérimentation.

L'expérimentation Découvrabilité et territoire

Objectif : Proposition d'un circuit de circulation automatique de la donnée d'événement, depuis les sites internet des lieux de diffusion vers le site de l'Office de Tourisme et des Congrès de la Métropole de Lyon.

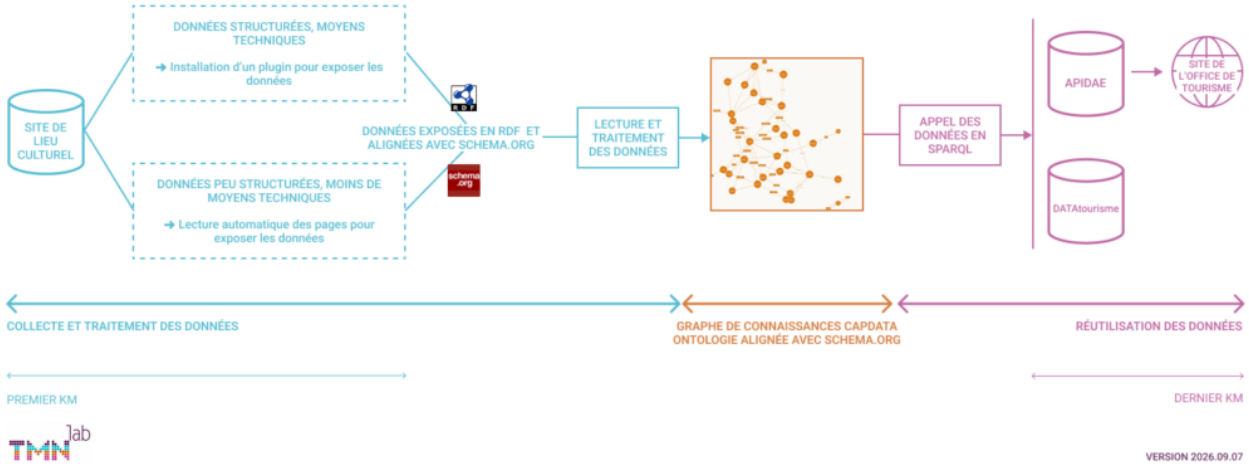
Enjeux :

- Réduction des saisies manuelles
- Diffusion d'une donnée fiable et à jour
- Proposition applicable à divers types de structures (via plusieurs modes de collecte des données)
- Modèle réplicable à plus large échelle

Parcours des données :

1. Exposition des données de la structure culturelle (alignement avec Schema.org)
    1. **Sites avec des données très structurées et des possibilités techniques** → Utilisation d'un plugin relié au CMS du site internet
    2. **Sites avec des données peu structurées ou moins de possibilités techniques** → Indexation automatique des contenus des pages de programmation du site Internet
2. Injection des données dans le graphe de connaissances CapData

3. Récupération des données utiles aux bases de tourisme par des requêtes en SPARQL du graphe de connaissances CapData
4. Affichage des données sur le site de l'Office de tourisme



Partenaires

- Opéra de Lyon
- ADN Tourisme (plateforme DATAtourisme)
- Apidae
- Solution Ideactiv
- Agence Logilab
- ONLYLYON Tourisme et Congrès, office de tourisme de la métropole de Lyon
- Agence Rezo Zero

D'autres structures culturelles ont accepté de participer aux tests comme le Périscope, l'Epicerie Moderne, le Théâtre Nouvelle Génération, le Théâtre de La Renaissance… Et d'autres progressivement.

Les premiers apprentissages

La nécessité d'avoir une URI par ressource décrite

Les graphes de connaissances sont constitués de triplets RDF, c'est-à-dire de descriptions de relations entre des informations à l'aide de 3 composantes et selon le langage RDF. Ces 3 composantes sont :

- l'identifiant de la ressource que l'on décrit
- la nature de la relation
- l'information de description

Les identifiants des ressources décrites sont des URIs (adresse vers la ressource sur le réseau).

L'enjeu est donc de fournir pour chaque élément décrit, de chaque événement de chaque lieu, une URI unique.

Identifiant unique de la source de la donnée

Lors de l'injection des données dans le graphe de connaissances, il faut pouvoir garder une référence à la source de la donnée (à la structure culturelle qui a produit la donnée). Pour cela, chaque lieu de diffusion doit pouvoir être identifié de façon unique et immédiate.

Pour cette première phase d'expérimentation, deux pistes ont été testées :

- S'appuyer sur le nom de domaine du site de la structure
- Se référer à la base des licences d'entrepreneurs de spectacle vivant

Les réflexions doivent être poursuivies pour aboutir à des options fiables, durables et surtout généralisables, qui pourraient à terme être utilisées plus globalement pour résoudre les problématiques de dédoublonnage (ex : un même spectacle accueilli et renseigné par plusieurs lieux différents).

Cela rejoint ainsi plus généralement tous les questionnements autour des stratégies de réconciliation de données, dès qu'il s'agit de réunir et de croiser des données (en amont, comme dans le graphe CapData, ou en aval comme dans DATAtourisme et Apidae). Pour fiabiliser, nettoyer et homogénéiser des données, comment peut-on automatiser une partie des traitements pour diminuer le poids des vérifications manuelles ?

De la description théorique à la donnée affichée

Plusieurs allers-retours ont été nécessaires pour fixer le périmètre et la structure des données de spectacle à exposer.

Les schémas à disposition sont des modèles de description sémantique et les données à afficher en sortie pour un usage spécifique sur une page web ne collent pas forcément de façon stricte à la structuration des données en entrée.

Très concrètement, pour l'affichage sur le site de l'office de tourisme de Lyon, il a fallu combiner les données de plusieurs classes Schema.org comme l'œuvre (CreativeWork), les informations de représentation (Event, Offer), les coordonnées de la salle de spectacle

(Organization, Place), etc.

Cela montre à la fois l'intérêt d'un modèle aussi développé que Schema.org, qui est assez étendu pour fournir des classes couvrant l'essentiel des besoins, et la nécessité de se doter de principes communs pour tout le spectacle vivant d'utilisation de ces classes.

Gestion des médias

Il y a d'une part la question des droits. Vidéos, photos, ressources PDF intégrant des visuels… Tous ces médias peuvent être soumis à des limites de diffusion liées aux droits négociés. Comment les respecter tout en permettant la diffusion des contenus sur d'autres plateformes ?

Plusieurs pistes de réponse ont été discutées. Pour cette expérimentation, c'est l'URL du média source (donc l'adresse du média sur le site internet du lieu) qui est diffusée avec la possibilité de préciser une date d'expiration et des informations sur les droits.

D'autre part, il y a la question des formats d'affichage, pour les images notamment. Si par exemple un théâtre expose ses visuels dans un format paysage adapté au design de son site Internet, comment s'assurer qu'ils s'afficheront correctement (non rognés, non pixellisés, etc.) sur une autre plateforme de diffusion qui utilise peut-être des vignettes en format portrait ?

La juste fréquence de mise à jour

Il s'agit de régler les différentes fréquences de rafraîchissement ou de lecture des données pour permettre d'afficher une donnée à jour en sortie sur le site de l'Office de tourisme même en cas de modification de dernière minute.

C'est un paramétrage qui concerne l'ensemble du parcours de la donnée car pouvant agir sur chaque point de lecture de donnée.

Plus les données sont lues et transmises régulièrement, plus il y a de chances d'avoir une information à jour, mais plus cela peut demander de ressources. Un juste équilibre est à trouver.

Ressources et pistes

Les ressources développées dans le cadre de Cap Data – France 2030

- **Plugins CapData d'exposition des données :**
  - SDK PHP : https://gitlab.com/capdataopera/capdataopera-php-sdk
  - WordPress *(à mettre à jour suite aux dernières évolutions)* : https://gitlab.com/capdataopera/capdataopera-wordpress-plugin
  - Drupal *(à mettre à jour suite aux dernières évolutions)* : https://gitlab.com/capdataopera/capdataopera-drupal-plugin
- **CapData :**
  - Documentation de la première version du projet : https://www.rof.fr/rof/capdata-opera.aspx

- Le projet sur GitLab : https://gitlab.com/capdataopera
- Entrepôt SPARQL : https://sparql.capdataopera.fr/

Pistes pour la suite de l'expérimentation

- Observations et retours d'expérience à mener sur la durée
- Choix des fréquences d'actualisation des données
- Maintenance et mises à jour de l'infrastructure (reprise des plugins *open source* par d'autres développeurs pour l'update et les évolutions, fonctionnement global et coûts)
- Appropriation de la structure d'exposition des données par d'autres acteurs
- Recherche autour de stratégies de réconciliation de données
- Évaluation de la valeur ajoutée réelle du dispositif
- Intégration d'autres structures culturelles, diversifiées, au test
- Ajout d'autres modes de collecte des données
- Ajout d'autres points de sortie des données avec des cas d'usage associés (via DATAtourisme, Data.gouv.fr, OpenAgenda…)

AGENDA
Une après-midi dédiée à la découvrabilité du spectacle vivant et au projet Cap Data – France 2030 se tiendra le 30 novembre 2026 à Paris lors des *Rencontres TMNlab* | *Médiation, attention, découvrabilité : nouvelles responsabilités culturelles à l'ère numérique*
Programme et inscription