



HAL
open science

Un graphe de données pour mesurer les Objectifs de Développement Durable et comprendre l'héritage des événements sportifs sur les villes

Wissal Benjira, Faten Atigui, Bénédicte Bucher, Malika Grim-Yefsah, Nicolas Travers

► To cite this version:

Wissal Benjira, Faten Atigui, Bénédicte Bucher, Malika Grim-Yefsah, Nicolas Travers. Un graphe de données pour mesurer les Objectifs de Développement Durable et comprendre l'héritage des événements sportifs sur les villes. 39ème Conférence sur la Gestion de Données – Principes, Technologies et Applications (BDA'23), Oct 2023, Montpellier, France. <hal-04601895>

HAL Id: hal-04601895

<https://hal.science/hal-04601895v1>

Submitted on 5 Jun 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

Un graphe de données pour mesurer les Objectifs de Développement Durable et comprendre l'héritage des événements sportifs sur les villes

Wissal Benjira*
DVRC, Pôle Léonard de Vinci
LASTIG, Univ. Gustave Eiffel, IGN
Paris La Défense, France
wissal.benjira@{devinci.fr,ign.fr}

Faten Atigui
CEDRIC, CNAM
Paris, France
faten.atigui@lecnam.net

Bénédicte Bucher
LASTIG, Univ. Gustave Eiffel, IGN
Saint Mandé, France
benedicte.bucher@ign.fr

Malika Grim-Yefsah
LASTIG, Univ. Gustave Eiffel, ENSG
Marne la Vallée, France
malika.grim-yefsah@ensg.eu

Nicolas Travers
DVRC, Pôle Léonard de Vinci
Paris La Défense, France
nicolas.travers@devinci.fr

ABSTRACT

La disponibilité croissante de données joue un rôle majeur dans le suivi du développement durable des villes. Produire des indicateurs mesurant les Objectifs de Développement Durable (ODD) est un enjeu sociétal, nécessitant des sources hétérogènes, pour établir des comparaisons spatio-temporelles. Notre travail se focalise particulièrement sur l'étude de l'impact des événements sportifs sur les indicateurs de Développement Durable en mettant en évidence les enjeux managériaux liés à l'organisation de tels événements. Nous proposons une approche reposant sur un framework simplifié pour évaluer l'héritage laissé par ces événements, en utilisant un lac sémantique représenté par une base de données graphe stockée dans Neo4j. Cette approche est illustrée par une étude de cas basée sur des données variées, permettant de comparer l'état du territoire hôte avant, pendant et après la Coupe d'Europe de Football 2016 et d'évaluer l'évolution d'indicateurs dans différentes zones géographiques.

ACM Reference Format:

Wissal Benjira, Faten Atigui, Bénédicte Bucher, Malika Grim-Yefsah, and Nicolas Travers. . Un graphe de données pour mesurer les Objectifs de Développement Durable et comprendre l'héritage des événements sportifs sur les villes. In *Proceedings of 39ème Conférence sur la Gestion de Données – Principes, Technologies et Applications (BDA'23)*.

1 CONTEXTE GÉNÉRAL

La disponibilité croissante de données couvrant des aspects variés de la réalité est une opportunité pour mieux observer et comprendre cette réalité dans sa complexité. Dans le domaine du développement durable urbain, ces données jouent un rôle crucial pour évaluer et mesurer l'impact des actions entreprises dans les villes [1, 2, 6]. Les Objectifs de Développement Durable (ODD) définis et reconnus par les pays membres des Nations Unies, mettent à disposition des indicateurs destinés à servir de fondement principal pour suivre les progrès vers la réalisation de ces ODD [8].

*Auteur correspondant

L'émergence d'événements sportifs, tels que les Jeux Olympiques, suscite un intérêt particulier en raison de leur influence significative sur les territoires hôtes. Ces événements sportifs sont des phénomènes spatio-temporels qui affectent structurellement, économiquement et socialement un territoire accueillant ces événements, générant ainsi un « héritage » [4, 5, 9]. L'étude de cet héritage repose sur la mesure comparative des différents impacts (ex. sociétaux, structurels, environnementaux, pratiques sportives, etc.) de ces événements à la fois sur une zone géographique et sur leurs évolutions au cours du temps.

L'étude de cet héritage sur les territoires nécessite différentes sources de données. Toutefois, la collecte des données peut s'avérer particulièrement complexe en raison de la diversité des sources, de la variété des formats, et finalement du manque de relations explicites entre les différentes données. L'examen de la littérature démontre une absence de cadre unificateur permettant de mobiliser des données pourvues d'hétérogénéités sémantiques.

Le sujet de la thèse s'intéresse plus précisément à la structuration de données et de métadonnées afin d'effectuer des analyses critiques et des comparaisons relatives à l'impact d'événements sur les espaces urbains. Ainsi, l'objectif principal est de fournir un framework de croisement de données pour élaborer et exploiter des indicateurs associés aux ODD. Ce sujet prend tout son intérêt pour les collectivités territoriales, pour des porteurs de projets numériques autour des pratiques sportives et pour les sponsors de grands événements.

2 MODÉLISATION PROPOSÉE

Dans la littérature, différentes approches de modélisation de schémas sont adaptées aux bases orientées documents ou colonnes, mais peu sur des bases de données graphes géolocalisées [7]. Ainsi, nous proposons une méthode de conception de lacs de données basée sur des bases de données graphes liés au contexte géomatique, i.e. intégrant des données géographiques, favorisant l'extraction et l'exploitation d'indicateurs ODD.

L'approche proposée repose sur un framework qui simplifie et facilite les différentes étapes d'évaluation de l'héritage en matière de développement durable; à savoir l'identification, le traitement et l'interprétation des données croisées pour obtenir des résultats plus



Figure 1: Sous-graphe Neo4j autour de l'événement sportif de la Coupe d'Europe de Football 2016

significatifs. Le lac sémantique obtenu est représenté et structuré sous forme de base de données graphe facilitant le rapprochement des données avec un schéma de graphe de données unifié, des règles d'enrichissement du graphe et des opérations sur le graphe pour produire des mesures et leur composition pour définir les indicateurs.

3 CAS D'ÉTUDE

Pour implémenter l'approche sémantique, nous développons une méthodologie de conception de base de données graphe guidée par les cas d'usages. Le prototype permet de valider et évaluer l'approche proposée. Nous avons choisi de considérer l'évaluation de l'impact de l'organisation d'événements sportifs sur l'indicateur du taux de chômage (commun à plusieurs ODD) en utilisant les différentes sources de données :

- Recensement de l'activité professionnelle par l'INSEE. La donnée est disponible à l'échelle des IRIS¹ et renseigne sur les caractéristiques des actifs, ex. le sexe, l'âge ou la catégorie socio-professionnelle.
- Géolocalisations des gares et stations par SmartIDF. Il s'agit d'un recensement des stations de transports en communs.
- Sites sportifs et installations sportives par SmartIDF. Il s'agit d'un recensement des installations sportives, ex. stades de football.

Pour illustrer, prenons l'événement « Coupe d'Europe de Football 2016 ». Nous avons pu établir le sous-graphe de la figure 1 offrant une vue globale des interconnexions entre les différentes entités.

Ce modèle de graphe facilite l'établissement d'indicateurs comparables dans le temps et dans l'espace, permettant ainsi des comparaisons du taux de chômage avant, pendant et après l'événement: l'héritage. Afin de manipuler le graphe et les indicateurs, le graphe est stocké dans Neo4j. Des requêtes Cypher interrogent aussi bien l'évolution temporelle du graphe que la géolocalisation du lieu.

¹IRIS - Îlots Regroupés pour l'Information Statistique

Cette géolocalisation est représentée à partir de sa description [3]. Par ailleurs, du fait que les indicateurs ODD reposent sur la composition de plusieurs mesures, notre Framework combine différentes requêtes pour calculer ces indicateurs de plus haut niveau.

Ainsi, l'évolution du graphe offre également la possibilité d'analyser ces indicateurs et d'observer les variations dans différentes zones géographiques. Les tendances démographiques, telles que les mouvements de population par sexe et par tranche d'âge, peuvent être analysées. De plus, l'analyse géospatiale permet d'identifier les zones où l'impact de l'événement est plus significatif. Les résultats obtenus peuvent être rapprochés à des cibles ODD et ainsi à des mesures d'héritages, permettant ainsi d'évaluer l'impact de l'événement par rapport à ces objectifs globaux de durabilité.

REFERENCES

- [1] K. Anderson, B. Ryan, W. Sonntag, A. Kavvada, and L. Friedl. 2017. Earth observation in service of the 2030 Agenda for Sustainable Development. *Geo-spatial Information Science* 20 (2017).
- [2] L. Ballerini and S. Bergh. 2021. Using citizen science data to monitor the Sustainable Development Goals: A bottom-up analysis. *Sustain. Sci* 16 (2021).
- [3] H. Chen, M. Vasardani, S. Winter, and M. Tomko. 2018. A Graph Database Model for Knowledge Extracted from Place Descriptions. *ISPRS International Journal of Geo-Information* 7 (2018).
- [4] M. Grim-Yefsah and B. Bucher. 2021. Towards Improving Knowledge Capitalization System for Sport Events Legacy. *Proceedings of the 11th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2019) - KMIS* (2021).
- [5] M. Harada. 2005. Hosting a mega-sports event and its impact upon city development challenges of the city of Osaka. Comparing Sports Policy, Sports Investment and Regional Development Initiatives in the Hosting of Sports Events in East Asia and Europe. (2005).
- [6] K. Kurowska, R. Marks-Bielska, S. Bielski, A. Aleknavicius, and C. Kowalczyk. 2021. Geographic Information Systems and the Sustainable Development of Rural Areas. *Land* 10 (2021).
- [7] J. Mali, S. Ahvar, F. Atigui, A. Azough, and N. Travers. 2022. A Global Model-Driven Denormalization. *Lecture Notes in Business Information Processing* 446 (2022).
- [8] United Nations. 2015. Transforming our World: The 2030 Agenda for Sustainable Development. (2015).
- [9] H. Preuss. 2019. Event legacy framework and measurement, *International Journal Sport Policy* Politics. *International Journal of Sport Policy and Politics* 11, 1 (2019), 103–118.