

 <p>Commission Règles et qualité</p>	<p>Groupe de Travail animé par :</p> 
<p>COMPTE-RENDU DE REUNION GT Ethique de l'usage de l'information et de la donnée géolocalisées</p>	
<p>Réunion du GT du 1^{er} décembre 2023 en visio-conférence de 14h30 à 16h30</p>	

Liste des participant(e)s

Présent(e)s : Agnès Ruffat (Région Occitanie), Marie Lambois (IGN), Fabien Castel (Murmuration), Marie Parent (Ville de Montréal), Serge Hombert (Métropole Européenne de Lille), Guillaume Chrétien (Cerema), François Robida (Robida Consultant), Richard Mongeau (Université Montréal), Pierre-Alban Bonin (FNCCR), Simon Gignac (Ville de Montréal, Grégory Delobelle (Région Centre Val de Loire), Bruno Thenail (Région Normandie), Pierre Pages (Toulouse Métropole), Hervé Halbout (HALBOUT Consultants), co-animateur, Anthéa Serafin (Ekitia), co-animatrice.

Compte-rendu synthétique

Lors de cette réunion, l'entreprise Murmuration a présenté aux autres participants 3 cas d'usage de données géolocalisées, visant à développer des indicateurs et des outils d'aide à la décision en matière environnementale.

Après une présentation de chaque cas d'usage, les participants ont débattu des enjeux éthiques que ceux-ci leurs évoquaient. Les points essentiels de ces discussions sont repris ci-dessous.

Cas d'usage n°1 : Tableau de bord environnemental pour l'aide à la prise de décision dans le domaine du tourisme

<https://murmuration-sas.com/en/homepage>

1. Contexte : Introduction de Murmuration : « *Comme toute activité économique et sociale, le tourisme a besoin de données et de recherches pour permettre et soutenir son développement réussi et durable. Pour contrôler la durabilité, il est nécessaire de disposer de données sur l'utilisation des ressources, telles que la terre, l'énergie et l'eau, les émissions et le traitement des déchets, la dégradation, la (sur)fréquentation, la congestion et les nuisances. L'évaluation des impacts sociaux, économiques et environnementaux permet de concevoir et de mettre en œuvre des politiques et des stratégies visant à maximiser la contribution positive du tourisme et à minimiser ses effets négatifs.* »
2. Production d'indicateurs environnementaux à partir de données d'observation de la Terre par satellite, dans plusieurs domaines :
 - a. Pollution de l'air
 - b. Pollution de l'eau (zone côtières, lacs)
 - c. Santé de la végétation
 - d. Artificialisation des sols
 - e. Ilots de chaleur urbaine
 - f. Changement climatique

3. Croisement avec des données de fréquentation touristique
 - a. Service « FluxVision » de « Orange Business Service »: statistiques anonymisées issues du réseau de téléphonie mobile. Donne le nombre de résidents et de touristes présent sur une zone, par tranche de 2H dans la journée ou par nuitée.
 - b. Les séries de données de fréquentation et d'indicateur environnementaux sont injectés dans des modèles d'IA pour identifier des tendances, des corrélations et isoler l'impact spécifique de l'activité touristique.
4. Intégration et présentation de ces indicateurs dans un tableau de bord interactif
 - a. Visualisation cartographique
 - b. Visualisation temporelle (courbes)
 - c. Mise à jour temps réel, historique dans le passé et projections dans le futur pour certains indicateurs (changement climatique)
 - d. Échelle locale au territoire ciblé (Commune, Communauté de Communes, Département, Région) et possiblement aussi plus large (nationale), pour pouvoir se comparer à des territoires voisins.
5. Enjeux éthiques identifiés :
 - a. Qualité des données représentant les flux de mobilité touristique.
 - b. Fiabilité des indicateurs produits pour mesurer l'impact environnemental du tourisme à l'échelle territoriale.
 - c. Formation des utilisateurs finaux concernant l'utilisation de cet outil, dans le cadre d'un processus décisionnel.
 - d. Contribution du projet au développement d'un tourisme durable.

Précisions sur le cas d'usage

Les outils développés par Murmuration ont pour objectif de prédire l'évolution de l'impact de certaines activités touristiques sur un territoire donné. Les modèles sont alimentés avec des données satellites, détaillées ci-dessus (disponibles pour tous les territoires, la plupart du temps gratuites, objectives, et qui permettent une vision historique depuis le lancement du satellite), ainsi qu'avec les données Flux Vision fournies par Orange. Orange n'ayant pas le monopole de la téléphonie mobile, les données fournies font l'objet d'un traitement préalable visant à les harmoniser et à les anonymiser. Malgré l'absence de traçabilité des traitements effectués sur les données brutes, Orange est le seul opérateur à proposer de telles données à cette échelle. Murmuration affine ses modèles tous les 6 mois pour actualiser les données.

Les destinataires de ces outils sont les institutionnels du tourisme (agences régionales, offices du tourisme ...), qui ne disposent pas de beaucoup de données environnementales, puisque jusqu'à récemment le tourisme était surtout appréhendé sous l'angle économique.

Enjeux éthiques discutés avec les participant(e)s du GT

1° Les producteurs de données/modèles devraient fournir aux utilisateurs de ceux-ci une information loyale concernant leur qualité, afin d'éviter toute mauvaise interprétation. Toute la chaîne de valorisation de la donnée devrait donc être attentive à la documentation de cette qualité et notamment renseigner les limites des données/modèles produits, ainsi que les utilisations non pertinentes. Concernant spécifiquement les modèles, les producteurs devraient définir et être transparent sur le seuil de précision jugé acceptable pour appuyer une décision.

2° Les utilisateurs (ici les institutionnels du tourisme) devraient eux-mêmes informer les destinataires finaux des incertitudes affectant les données sur lesquelles ils fondent leurs décisions.

3° La formation de ces utilisateurs à la bonne compréhension des outils mis à leur disposition serait très pertinente. Pour dispenser une telle formation, le fournisseur de l'outil semble être l'acteur le mieux placé. Au-delà de cette formation continue dans le milieu professionnel, une pédagogie dans le milieu scolaire/supérieur concernant les enjeux liés à la fiabilité de l'information apparaît complémentaire.

4° [Il serait intéressant que chaque acteur de la chaîne de valorisation des données nomme un garant de l'usage éthique des données](#), à qui toute personne intéressée pourrait s'adresser pour se renseigner sur la qualité des données modèles, les conditions optimales d'utilisation d'un outil, ...

Cas d'usage n°2 : [UVEG (University of Valencia)] Migrations induites par le climat en Afrique

José María Tárraga, Maria Piles, Gustau Camps-Valls, Learning drivers of climate-induced human migrations with Gaussian processes, NeurIPS 2020 Workshop on Machine Learning for the Developing World <https://doi.org/10.48550/arXiv.2011.08901>

- a. Objectifs
 - i. Modéliser, anticiper et comprendre les flux migratoires induits par le climat (vagues de chaleur, sécheresses et inondations) en Afrique, à partir de données fiables.
 - ii. Dévoiler les dynamiques existantes entre toutes les variables considérées, grâce à un modèle d'inférence causale, afin de comprendre, quantifier et prévoir les effets des flux migratoires à partir du contexte socio-économique et de variables environnementales extraites de données d'observation de la Terre.
 - iii. Utilisateurs ciblés : agences nationales et inter-gouvernementales telles que l'iDMC (« Internal Displacement Monitoring Centre ») et le WFP (« World Food Program »)
- b. Données
 - i. Conditions météorologiques : températures, précipitations, vent
 - ii. Observations du satellite MODIS : capteur multispectral dans le visible et l'infrarouge, utilisé pour mesurer la santé de la végétation et l'artificialisation des sols
 - iii. Données socio-économiques : prix du marché local, conflits violents, pandémies, achat/vente de bétail...
- c. [Enjeux éthiques identifiés => analyse non effectuée](#)

Précisions sur le cas d'usage

Les données socio-économiques utilisées par les chercheurs pour alimenter leurs modèles sont de sources très variées et il est notamment possible d'y trouver des données en provenance de réseaux sociaux.

Des recherches ont par ailleurs été menées concernant la définition d'un indice de causalité probable, issu de la corrélation entre deux variables.

Enjeux éthiques discutés avec les participant(e)s du GT

1° Les producteurs des modèles devraient être attentifs à la qualité des données socio-économiques utilisées et la documenter (par exemple en accompagnant leurs publications scientifiques de *data papers*).

2° Les producteurs des modèles devraient informer les utilisateurs finaux des hypothèses à l'origine de liens de causalité supposés entre deux variables.

3° Les conditions d'utilisation des modèles devraient être prédéfinies, notamment afin d'éviter toute récupération à des fins politiques, voire démagogiques, des données sortantes. Il serait pour cela important de définir une typologie de personnes habilitées à les utiliser et d'encadrer à quelles fins elles peuvent être utilisées.

4° Afin de faciliter la réutilisation des résultats de leurs recherches, les chercheurs peuvent par exemple s'appuyer sur les principes FAIR (*Findable/Facile à trouver, Accessible/Accessible, Interoperable/Interopérable, Reusable/Réutilisable*). Il conviendrait toutefois de préciser les points de vigilance à prendre en considération concernant l'application de ces modèles à d'autres zones géographiques.

5° Les utilisateurs finaux des données sortantes des modèles devraient être transparents (à l'égard des destinataires de l'information fournie sur la base de ces données) sur le niveau d'incertitude affectant celles-ci.

Cas d'usage n°3 : [National Observatory of Athens] Prévission à court terme des risques d'incendie en Méditerranée

Ioannis Prapas, Spyros Kondylatos, Ioannis Papoutsis, Gustau Camps-Valls, Michele Ronco, Miguel-Ángel Fernández-Torres, Maria Piles Guillem, Nuno Carvalhais, Deep Learning Methods for Daily Wildfire Danger Forecasting, <https://doi.org/10.48550/arXiv.2111.02736>

a. Objectifs

- i. Améliorer la planification opérationnelle, la préparation et la capacité d'intervention en cas d'urgence, des pompiers et d'autres décideurs (par exemple, les autorités de protection civile, les ministères de l'environnement, etc.), en leur fournissant un système de prévission des risques d'incendie robuste et fondé sur des données probantes, avec une résolution spatiale et une précision supérieures à celles de l'état de l'art.
- ii. Pour répondre aux besoins de nos utilisateurs finaux, trois produits de base sont proposés :
 1. Produit 1 - Prévissions quotidiennes du danger d'incendie
 2. Produit 2 - Graphiques d'explicabilité
 3. Produit 3 - Évaluation quotidienne du risque d'incendie

b. Données

- i. Conditions météorologiques : températures, précipitations, vent
- ii. Observations du satellite MODIS : capteur multispectral dans le visible et l'infrarouge, qui permet notamment de mesurer la santé de la végétation
- iii. Données d'humidité dans le sol produites par le JRC (« Joint Research Center », laboratoire de recherche de la Commission Européen)
- iv. Données statiques environnementales : topographie, couverture des sols, densité de population, réseau routier, surfaces en eaux (lacs, rivières...)

c. Méthodologie

- i. Combiner l'utilisation de modèles physiques et d'un modèle de « Deep Learning » basé sur des données historiques de feux de forêt en Grèce.
- ii. L'objectif est d'identifier les facteurs influant le déclenchement d'un feu, pour ensuite produire un modèle de prédiction qui fournit un indice quotidien de risque de feu.

d. Enjeux éthiques identifiés => *analyse non effectuée*

Précisions sur le cas d'usage

Cet outil a été utilisé par les pompiers grecs en 2022 et 95% des incendies sont survenus dans les zones indiquées à fort risque par l'outil. À noter toutefois qu'il y également eu beaucoup de faux positifs. Les Services Départementaux d'Incendies et de Secours sont pour l'instant habitué à utiliser des plans dotés de SIG, mais ne sont pas encore familier des applications prévisionnelles.

Enjeux éthiques discutés avec les participant(e)s du GT

1° Les producteurs des modèles devraient fournir aux utilisateurs finaux une information loyale concernant les limites des données utilisées pour les alimenter, ainsi que sur leurs conditions optimales d'utilisation (par exemple concernant le biais lié à l'alimentation du modèle, uniquement avec des données historiques).

2° Les conditions d'utilisation et de réutilisation des données sortantes du modèles devraient être bien encadrées et prédéfinies (limitées à un objectif strict de prévention des incendies : il serait en effet discutable qu'elles puissent être réutilisées par des promoteurs immobiliers ou des assureurs, par exemple)

3° Il existe une incertitude concernant la communication à la population d'une information liée à une probabilité d'incendie. Les acteurs publics sont soumis à une obligation de communication des informations environnementales qu'ils détiennent, ainsi que d'ouverture de certaines données publiques. Leur responsabilité peut donc être mise en cause en cas de manquement à ces devoirs. Toutefois éthiquement, une question se pose concernant la communication d'informations peu fiables / peu probables. Il ne serait en effet peut-être pas bienvenu qu'une prédiction concernant un très faible risque d'incendie entraîne un mouvement de panique au sein de la population. Il pourrait par ailleurs être reproché aux autorités de ne pas être intervenues, alors qu'elles détenaient une information utilisable, même si celle-ci relèverait d'une faible probabilité de survenance d'incendie. Sur cette question de la responsabilité, un participant au GT rappelle une situation similaire s'étant produite dans le domaine sismique : après une série de séismes survenus en Italie, des experts avaient été consultés pour avis sur la probabilité de répliques et, sur la base des données qu'ils détenaient, ils se sont positionnés en faveur d'un faible risque de répliques. Les répliques se sont en réalité avérées importantes et certains experts ont été condamnés à de la prison. L'idée à retenir de cette comparaison est qu'il faudrait éviter que, par craintes de retombées négatives, des chercheurs ou des autorités procèdent à une rétention de certaines informations.

Remarques générales

L'enjeu de la documentation et de la transparence concernant la qualité et la fiabilité des données géolocalisées / modèles fondés sur des données géolocalisées, par tous les acteurs de la chaîne de valorisation (du producteur à l'utilisateur final) ressort dans les 3 cas d'usage présentés. Il semble en effet essentiel que les utilisateurs finaux prenant des décisions sur la base de données, comprennent les limites de celles-ci, soient transparent à cet égard vis-à-vis des personnes impactées par leurs décisions, tout en assumant leur niveau de responsabilité quant aux conséquences de celles-ci.