



---

**Mesurer la structuration de données  
au service de la découvrabilité en ligne**  
Résultats du projet dia-log

---

août  
2023

dia-  
-log

Gradiant a eu l'honneur d'être sollicité par dia-log pour la mise en place et le suivi d'indicateurs de performance dédiés à leur initiative. Ce document représente le bilan de notre mandat et détaille les indicateurs élaborés, la méthodologie adoptée, ainsi que les principaux résultats obtenus. Nous y apportons également un regard sur les perspectives d'amélioration et les limites de notre étude.

Né de la vision du Réseau des conseils régionaux de la culture du Québec, dia-log est un projet ambitieux s'échelonnant d'avril 2020 à septembre 2023 qui visait à sensibiliser les acteurs de la diffusion et de la promotion des arts de la scène aux enjeux de la découvrabilité en ligne. Son objectif principal était de déployer des mesures concrètes pour améliorer la qualité, l'ouverture et le potentiel de réutilisation des métadonnées des événements en arts de la scène. Enrichi par la collaboration de nombreux experts et partenaires, notamment des domaines juridique et technologique, dia-log a représenté une avancée significative pour le milieu des arts de la scène au Québec.

# Table des matières

<b>Sommaire exécutif.....</b>	<b>1</b>
<b>Introduction.....</b>	<b>2</b>
Qu'est-ce que la découvrabilité ?.....	2
Qu'est-ce que la structuration des données ? Et à quoi sert-elle ?.....	2
Pourquoi mesurer la découvrabilité ?.....	2
<b>Méthodologie.....</b>	<b>3</b>
Quelques mots sur les indices de mesure de découvrabilité existants.....	3
Privilégier l'impact dans le déploiement des indicateurs.....	3
Nos indicateurs et les dimensions d'analyse.....	3
Les objectifs par type de contributeur.....	3
Les mesures observées.....	5
Les dimensions observées.....	6
Nos données.....	6
Qu'avons-nous analysé et comment l'interpréter ?.....	7
Limites de l'étude et notes d'interprétation.....	7
<b>Quels résultats ?.....</b>	<b>8</b>
Présence et capture : donner de l'information pratique pertinente à ses usagers.....	8
Visibilité : Visibiliser les événements disponibles sur son territoire.....	8
Consultation : Développer ses publics.....	8
Des observations différentes selon les caractéristiques des contributeurs.....	10
Des observations rythmées par la saisonnalité.....	11
Comprendre les liens entre la durée de mise en ligne et le nombre de pages vues.....	12
<b>Conclusion.....</b>	<b>14</b>
Perspectives d'amélioration.....	14
<b>Références.....</b>	<b>15</b>
<b>Annexes.....</b>	<b>16</b>
Annexe 1. Schéma de transformation des données.....	17
Annexe 2. Limites des données à disposition.....	18
Annexe 3. Définitions d'un événement structuré, non structuré et semi-structuré.....	19

# Sommaire exécutif

**Contexte :** La découvrabilité dans le domaine culturel est essentielle pour maximiser la visibilité et l'impact des événements culturels en ligne. La plupart des efforts pour mesurer la découvrabilité ont été axés sur l'évaluation des bonnes pratiques, laissant un écart dans l'évaluation de l'impact réel de ces pratiques sur la découvrabilité, et de son retour sur investissement pour les organismes.

**Objectif :** Notre étude vise à combler cet écart en développant des indicateurs pouvant être réutilisés dans des analyses subséquentes, d'une part, et en analysant ces derniers dans le contexte du projet dia-log.

**Méthodologie :** Particulièrement, nos indicateurs visent à évaluer la présence et la capture de nouvelles informations pertinentes pour les usagers, le potentiel de réutilisation des données structurées et la consultation des pages d'événements (intérêt du public) selon qu'ils étaient structurés ou non. Nous avons, par ailleurs, observé ces indicateurs selon un ensemble de caractéristiques liées au contributeur et les caractéristiques des événements.

## Résultats :

- La structuration de données a permis une augmentation significative de la pertinence de l'information offerte aux usagers : 3 251 événements ont été structurés, liés à 54 lieux et 212 artistes.
- Les événements dont les données ont été structurées ont enregistré un nombre de pages vues moyen supérieur de 17 % aux événements dont les données n'étaient pas structurées, traduisant une meilleure interaction avec les internautes.
- L'augmentation de la consultation moyenne de pages demeure inégale d'un contributeur à un autre : les diffuseurs et les organismes qui n'avaient pas du tout de données structurées préalablement au projet sont ceux qui ont le plus bénéficié de la structuration de leurs données.
- La structuration de données ne semble avoir d'effet que pour un événement dont la période de mise en vente est suffisante (supérieure à 3 mois).

Notre étude a des limites. L'absence d'un groupe de contrôle peut affecter la validité de nos conclusions. De plus, des facteurs tels que la saisonnalité, les biais dans les données, et les erreurs potentielles dans les bases de données peuvent influencer nos résultats.

Nos indicateurs offrent cependant une nouvelle manière d'analyser l'impact et l'efficacité de la structuration et la mutualisation des métadonnées. Tout en reconnaissant les limites de notre étude, nous croyons que nos résultats fourniront des informations précieuses pour celles et ceux qui cherchent à améliorer la visibilité et l'impact de leurs événements culturels en ligne.

# Introduction

## Qu'est-ce que la découvrabilité ?

La découvrabilité désigne la capacité d'un produit ou d'un contenu à être découvert, particulièrement dans le contexte d'un paysage numérique vaste et saturé. C'est un enjeu essentiel pour le secteur culturel où le rayonnement est crucial pour le succès d'un événement.

Rendre son contenu ou son offre découvrable implique la mise en place d'une stratégie alliant plusieurs méthodes, allant de l'optimisation pour les moteurs de recherche (SEO) ou l'utilisation efficace des médias sociaux à la mise en œuvre de tactiques de marketing et de promotion ciblées. Le choix des méthodes et de la stratégie de découvrabilité en place sera donc intimement lié aux objectifs poursuivis par l'organisme à l'égard de ses publics. La structuration de ses données est l'une des méthodes préconisées.

## Qu'est-ce que la structuration des données ? Et à quoi sert-elle ?

Une donnée est une information brute qui peut être collectée et analysée. Cette information peut se présenter sous divers formats, tels que du texte, des chiffres, etc. La diversité de ces formats peut compliquer leur exploitation et leur compréhension. C'est là qu'il devient alors nécessaire de les structurer.

Structurer ses données signifie organiser ces informations de manière à les rendre uniformément compréhensibles et exploitables. C'est l'équivalent de l'adoption de vocabulaires ou de règles de grammaire dans une langue afin de pouvoir se comprendre.

En ce qui concerne un site Web, structurer ses données se fait, par exemple, par l'adoption de vocabulaires utilisés par les créateurs de site Web ou les moteurs de recherche (p. ex. balises schéma). Utiliser ces vocabulaires aide non seulement les moteurs de recherche à mieux comprendre le contenu d'un site Web, mais aussi des initiateurs externes à créer

toutes sortes d'applications (p. ex. La Vitrine).

## Pourquoi mesurer la découvrabilité ?

La mise en place d'une stratégie de découvrabilité, notamment la structuration de données, représente un investissement considérable. Malgré le soutien gouvernemental, le milieu culturel québécois perçoit ces efforts comme coûteux et chronophages. Conséquemment, les méthodes liées directement aux plateformes, telles que les algorithmes et la taxonomie, viennent loin derrière les méthodes traditionnelles. (Wells & Desjardins, 2020).

Ces éléments soulignent la nécessité de mesurer l'impact de la structuration de données pour évaluer leur rentabilité. Jusqu'à présent, les mesures de découvrabilité ont principalement visé à évaluer la découvrabilité des contenus québécois dans leur ensemble plutôt que d'aider des organismes spécifiques à imaginer les stratégies adéquates afin de faire rayonner leurs contenus. **Cette étude vise à mettre en lumière l'impact de la structuration de données, en plus de servir de cap aux organismes.**

# Méthodologie

## Quelques mots sur les indices de mesure de découvrabilité existants

Dans le cadre de notre veille, deux familles de mesures ont été identifiées.

- **Les mesures de type marqueurs de rayonnement** sont les mesures qui cherchent à représenter explicitement le rayonnement.
- **Les mesures de type indices sont les mesures qui cherchent à représenter les facteurs de rayonnement** - c'est-à-dire, dans quelle mesure les meilleures pratiques favorisant le rayonnement ont-elles été appliquées.

Les études les plus abouties portent davantage sur les indices. Peu d'efforts ont été déployés afin d'évaluer la

découvrabilité à proprement parler et l'impact que ces différentes pratiques pourraient avoir sur celle-ci.

## Privilégier l'impact dans le déploiement des indicateurs

Puisque la majorité des indicateurs mesurent les facteurs influençant le rayonnement, nous avons décidé de non seulement étudier la découvrabilité sous ce prisme, mais également sous le prisme de l'impact réel sur le rayonnement.

La méthodologie s'approche de l'action 7 des propositions d'actions fournies dans l'Étude de faisabilité pour la conception d'indicateurs de rayonnement de l'offre culturelle québécoise dans les réseaux numériques de l'Institut de la Statistique du Québec (2021). Celle-ci consiste à évaluer les effets de l'enrichissement et de la structuration des données d'un panier d'offres culturelles sur leur rayonnement sur le Web.

## Nos indicateurs et les dimensions d'analyse

Afin de déterminer quoi mesurer, il faut d'abord savoir pourquoi nous souhaitons accroître notre découvrabilité. Nous avons donc d'abord inventorié les objectifs poursuivis par les organismes lorsqu'ils mettent en ligne des événements sur leurs sites Web. Les mesures observées ont ensuite été développées pour suivre ces objectifs.

### Les objectifs par type de contributeur

Deux types de contributeurs ont participé au projet dia-log : des diffuseurs et des agrégateurs. Les agrégateurs sont, par exemple, des villes ou des regroupements qui ont créé un calendrier agrégé des événements ayant lieu sur leur territoire ou organisés par leurs membres.

Ainsi, voici une liste non exhaustive des objectifs poursuivis par ces deux types d'organismes :

#### 1. Objectifs d'un agrégateur

- Visibiliser les événements disponibles sur son territoire
- Donner de l'information pratique pertinente aux citoyens et usagers de leur site Web : le lieu de l'événement, les informations sur l'artiste, etc.

## **2. Objectifs d'un diffuseur**

- Développer ses publics
- Donner de l'information pratique à ses publics

## Les mesures observées

Objectifs	Mesures	Indicateurs	Interprétation
<b>Donner de l'information pratique pertinente à ses usagers : le lieu de l'événement, les informations sur l'artiste</b>	Présence et capture de nouvelles entités	Nombre d'événements capturés	La présence permet d'évaluer les événements déjà existants "dans le Web", c'est-à-dire si leurs données sont structurées afin de pouvoir éventuellement être lus. Ce nombre permet simplement d'obtenir un ordre de grandeur. Les événements ont été comptabilisés sur le graphe de connaissances Artsdata.ca et à travers un moissonnage (certains contributeurs n'avaient pas encore leurs données entrées dans Artsdata.ca).
		Pourcentage des événements avec des artistes, lieux, liens de billetterie	Plus ce pourcentage est élevé et proche de 100 %, plus les événements ont un degré de données structurées élevé. Le pourcentage pour les lieux, les noms et les dates sera toujours à 100 %, tandis que les artistes, les liens URL vers la billetterie ou la description sont des informations facultatives.
<b>Visibiliser les événements disponibles sur son territoire</b>	Potentiel de réutilisation des données structurées	Nombre de liens potentiels vers d'autres plateformes et bases de connaissances par artiste	La visibilité fait normalement référence au fait qu'un événement soit relayé ou non par d'autres instances (médias, réseaux sociaux, plateformes). Ici, nous avons plutôt observé le nombre de liens potentiels qui peuvent être faits avec d'autres plateformes et bases de connaissances, à partir du moment où les événements sont collectés sur la plateforme Artsdata.ca. Par exemple, ces plateformes pourraient être Facebook, Wikidata ou encore MusicBrainz. Plus ce nombre est élevé, plus il y a de potentiel d'interopérabilité et donc de visibilité.
<b>Développer ses publics</b>	Consultation de l'offre culturelle	Nombre de pages vues en moyenne par événement par jour	Le nombre de fois où une page d'événement a été vue est une information approximative quant à l'intérêt que peut porter les publics à l'égard de cet événement. Plus ce nombre est grand, mieux c'est.
		Nombre de pages vues en moyenne par type d'événement	Puisque tous les événements ne sont pas toujours mis en ligne pendant la même durée, et que nous n'avons pas toujours tout l'historique des données de navigation sur les pages, nous préférons observer le nombre de pages vues moyen par événement par jour.



## Les dimensions observées

Nous avons également observé les résultats sous le prisme de plusieurs dimensions afin d'en tirer des spécificités. Ainsi, nous les avons étudié également selon le profil des contributeurs, c'est-à-dire, selon :

- Qu'ils ont effectué une refonte de leur site Web ou non dans les 5 dernières années;
- Qu'ils sont agrégateurs ou diffuseurs;
- Qu'ils diffusent ou non leurs événements sur d'autres plateformes (p. ex Facebook);
- La méthode de structuration de données choisie (JSON-LD ou l'outil Footlight développé par La Culture Créée);
- Leur nombre d'années d'existence;
- Leur niveau de données structurées au début du projet :
  - Inexistant;
  - Minimal;
  - Appréciable;
  - Élevé;
- Leur taille en termes de volume d'événements en arts de la scène par année.

## Nos données

Trois bases de données (BD) ont été créées afin de construire ces mesures.

- **BD Événements** : recense la liste des événements dont les données ont été structurées grâce à 1) des requêtes SPARQL spécifiques sur Artsdata.ca (pour la majorité des contributeurs) afin de collecter les événements structurés à venir, 2) des extractions des données des événements structurés entreposées sur Amazon S3 par La Culture Créée et 3) le moissonnage des données à même le site Web de certains contributeurs.
- **BD Artistes** : recense la liste des artistes présents dans Artsdata.ca grâce à des requêtes SPARQL spécifiques.
- **BD Consultation** : recense le nombre de pages vues par page d'événement par jour grâce à l'extraction des données de consommation et de consultation de pages d'événements, qui a été réalisée manuellement en téléchargeant les données en .xlsx depuis l'interface Google Analytics. Toutefois, on pourrait tout à fait imaginer effectuer des requêtes sur le Google Analytics API pour être plus efficace à l'avenir. Nous y retrouvons également les caractéristiques des contributeurs, essentiellement obtenues via des questionnaires.

Un schéma de la transformation des données, disponible en annexe 1, présente comment nous avons obtenu ces bases de données. Les données à disposition présentent des limites, qui ont été listées en annexe 2.

## **Qu'avons-nous analysé et comment l'interpréter ?**

Nos indicateurs et les analyses qui en sont faites cherchent à répondre aux questions suivantes :

1. Quel est le niveau de présence des événements sur le Web ?
2. Quel est le niveau de visibilité et de notoriété des contributeurs et des spectacles sur d'autres plateformes ?
3. Quel est le niveau de consommation (consultation) de notre contenu ?
4. Dans quelle mesure la structuration des données a-t-elle permis d'étendre le rayonnement (la consultation) des contenus culturels ?

## **Limites de l'étude et notes d'interprétation**

Afin de répondre à la dernière question portant sur les effets de la structuration sur le rayonnement, il est nécessaire de mettre en place un test A/B, i.e. de comparer un panier d'événements dont les données seraient structurées (groupe de test) à un autre dont les données ne le seraient pas (groupe de contrôle). Compte tenu de la complexité que cela demandait pour les organismes et leurs fournisseurs, cela n'a pas pu être mis en place. Étant donné que les événements n'ont pas pu être sélectionnés de manière aléatoire et en quantité suffisante, les résultats que nous allons présenter par la suite doivent être interprétés comme des corrélations et non de la causalité. Grâce aux rapports individuels partagés avec chacun des contributeurs, ces derniers pourront toutefois accéder aux résultats et les interpréter selon leurs propres connaissances des événements et de leurs contextes respectifs.

## Quels résultats ?

### **Présence et capture : donner de l'information pratique pertinente à ses usagers**

Le projet dia-log **a permis de structurer les données de 3 251 événements**. Ces événements ont été associés à 54 lieux et 212 artistes référencés dans Artsdata.ca ou Wikidata.

Il reste que seulement 33 % des événements sont associés à un artiste référencé, ce qui suppose qu'il devrait y avoir un effort complémentaire de la part des producteurs, artistes et/ou celles et ceux qui les représentent afin de les ajouter sur ces plateformes. De plus, 55 % des événements avaient des informations de billetterie (c.-à-d. lien URL, prix).

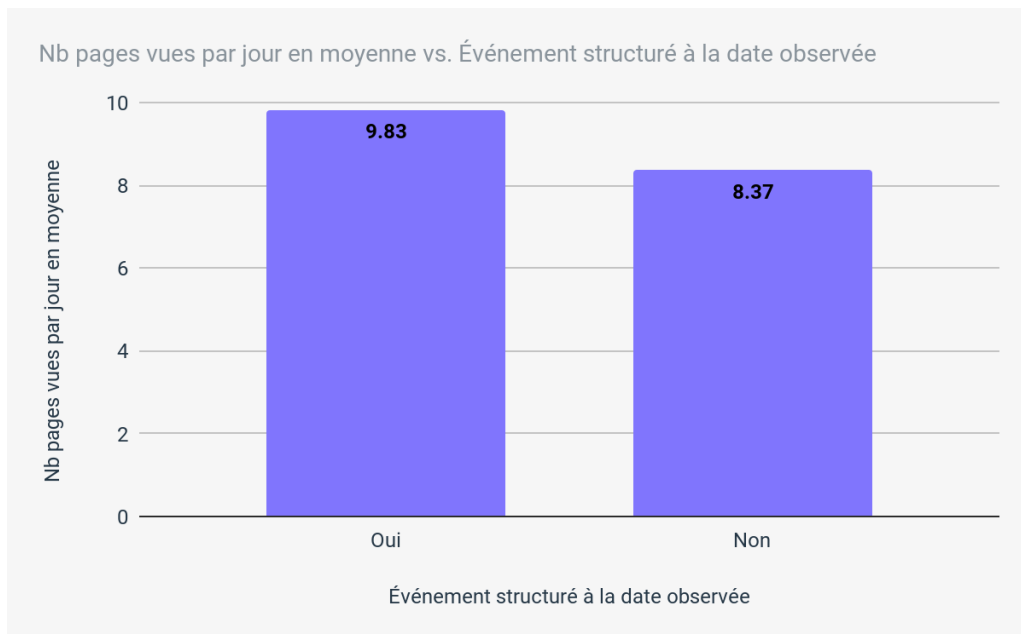
Les diffuseurs ont présenté des pourcentages de capture d'artistes plus élevés que les agrégateurs. De plus, les organismes en région ont montré des pourcentages de capture d'artistes plus élevés que ceux en métropole. Cela suggère que les organismes en ville ressentent moins le besoin de structurer leurs données puisqu'ils bénéficient de plus de visibilité en raison de leur situation géographique.

### **Visibilité : Visibiliser les événements disponibles sur son territoire**

En moyenne, les artistes identifiés dans les événements structurés dans le cadre de dia-log ont près de 3 liens enregistrés vers d'autres plateformes. Les organismes en métropole affichaient un nombre moyen de liens vers d'autres plateformes plus élevé que ceux en région, posant la question de la notoriété et/ou de l'importance de ces liens par les artistes se produisant en métropole.

### **Consultation : Développer ses publics**

La structuration de données **semble avoir un effet positif sur le nombre de pages vues**.

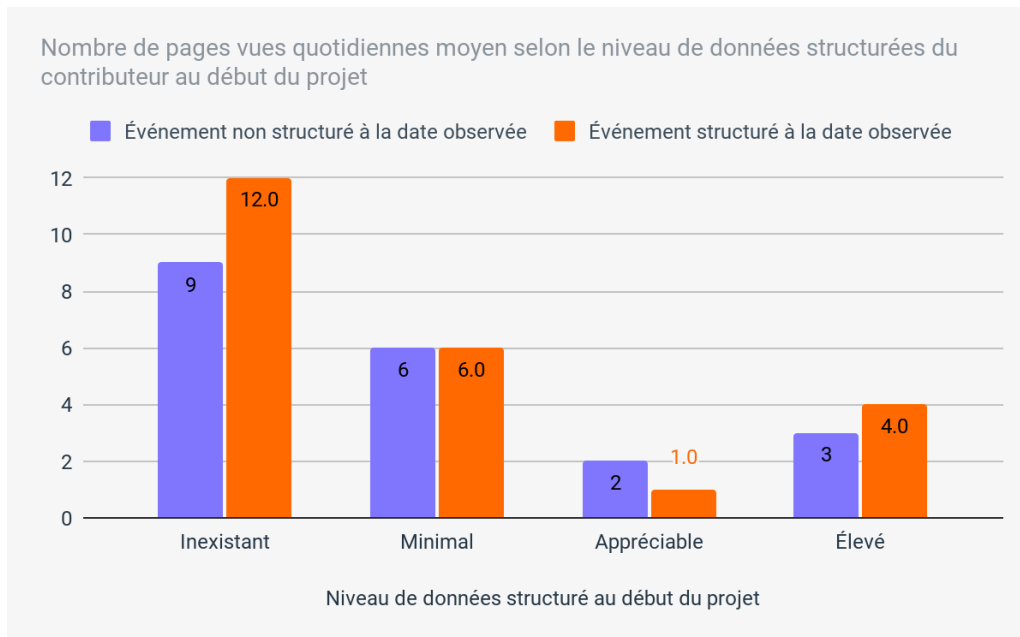


**Le nombre moyen de pages vues quotidiennes moyen structuré est 17.4 % supérieur à celui d'un jour non structuré.**

Un événement partiellement structuré<sup>1</sup> a également observé plus de quatre fois plus de pages vues quotidiennes en moyenne par rapport à un événement non structuré.

<sup>1</sup> Définitions d'événements structurés, partiellement structurés et non structurés en annexe 3.

## Des observations différentes selon les caractéristiques des contributeurs

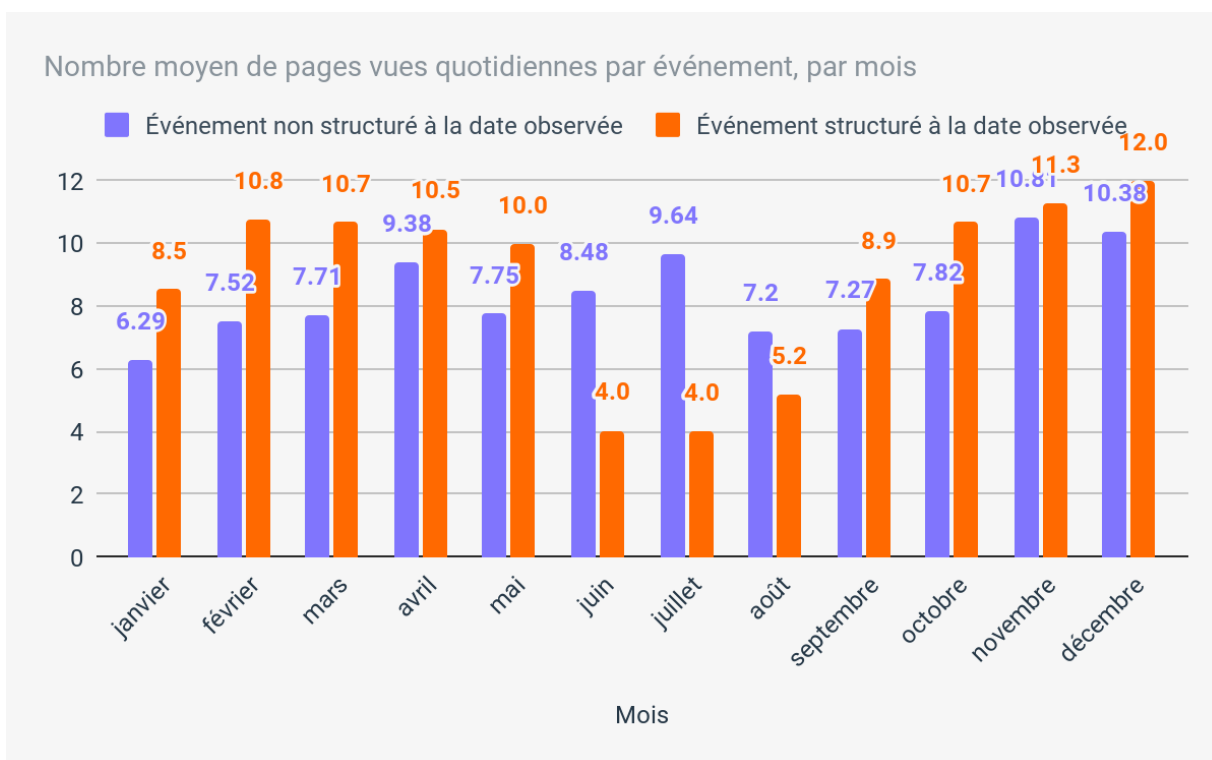


Cette augmentation était surtout observable pour les événements dont le **niveau de données structurées initial était inexistant**, pour les **diffuseurs** et les **contributeurs ayant opté pour la méthode de structuration de données JSON-LD**.

Les contributeurs n'ayant pas fait de refonte de leur site Web dans les 5 dernières années ont également connu une augmentation, avec **une moyenne de 30 % de pages vues quotidiennes de plus** par événement lors d'un jour structuré comparativement à un jour non structuré.

## Des observations rythmées par la saisonnalité

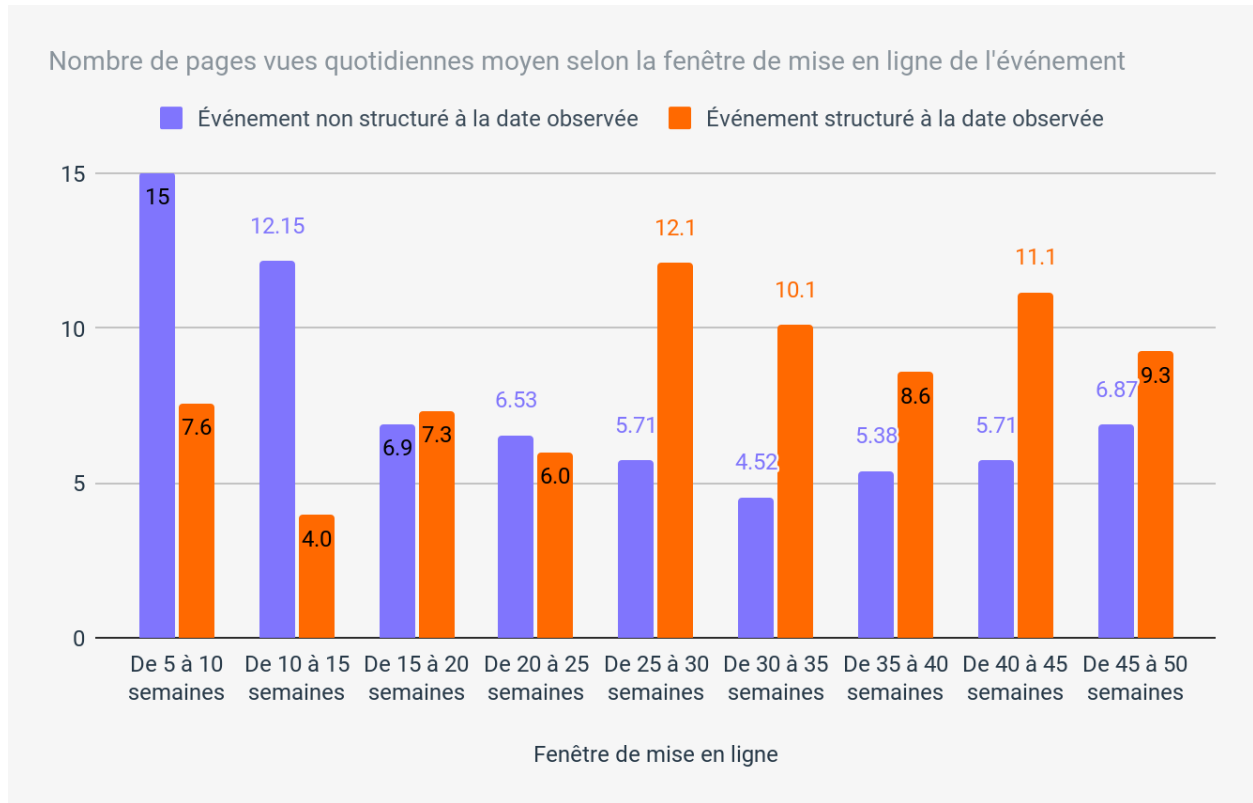
Afin d'analyser les variations saisonnières, nous avons évalué la distribution mensuelle.



Il ressort que la période de janvier à mai enregistre une hausse significative, **avec une augmentation d'environ 40 % entre un jour avec des données structurées et un sans**. Notons que la période de septembre à décembre affiche un volume de pages vues quotidien naturellement élevé, probablement dû aux lancements saisonniers, qui pourrait expliquer un impact moins important des données structurées.

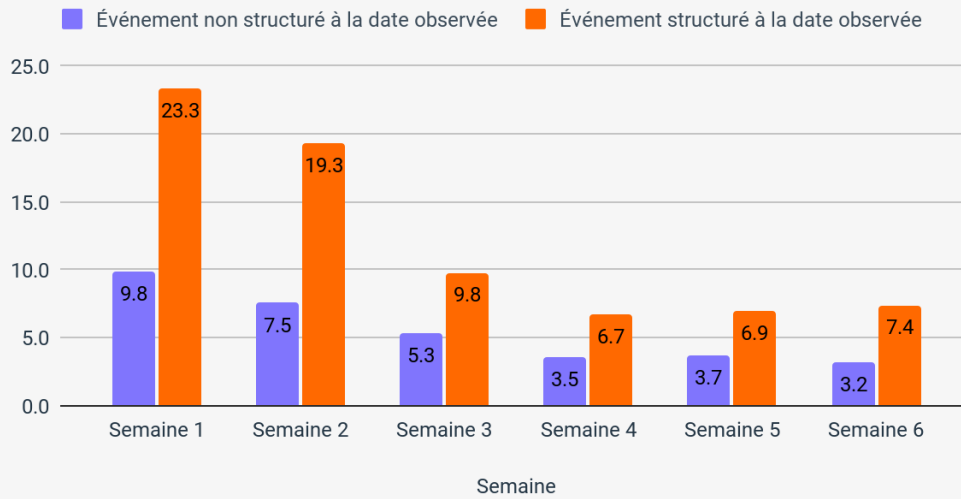
## Comprendre les liens entre la durée de mise en ligne et le nombre de pages vues

L'examen de l'influence de la fenêtre de mise en ligne des événements était également crucial.



Les événements bénéficiant d'une fenêtre de mise en ligne inférieure à 15 semaines montrent un nombre réduit de pages vues quotidiennes lorsqu'ils sont structurés. Ceci pourrait être associé à une première semaine disproportionnée ou à des fenêtres de mise en ligne incomplètes. À l'opposé, **les événements ayant une fenêtre de mise en ligne s'étendant de 25 à 50 semaines enregistrent, en moyenne, un taux de pages vues quotidien plus élevé lorsqu'ils sont structurés.**

Nombre de pages vues quotidiennes en moyenne dans les 6 premières semaines pour les événements dont la fenêtre de mise en ligne est de 25 à 30 semaines



En particulier, les événements ayant entre 25 et 30 semaines de mise en vente présentent un volume de pages vues **plus de 2 fois supérieur lorsque structurés** dans les 6 premières semaines de mise en ligne, comparativement à lorsqu'ils ne le sont pas.



# Conclusion

Les événements dont les données ont été structurées ont globalement observé un plus grand nombre moyen de pages vues quotidiennes que les événements dont les données n'étaient pas structurées. Malgré cela, ces tendances varient grandement selon le contributeur.

Cela réitère le besoin d'inscrire la structuration des données dans une stratégie globale de découvrabilité. Il semble judicieux d'entreprendre cette tâche après avoir consolidé les autres piliers stratégiques, car des facteurs tels que la notoriété ou la taille du contributeur semblent déjà influencer considérablement la visibilité des événements qu'il présente ou relaie. Les résultats nous indiquent aussi que ce sont les diffuseurs qui devraient porter une attention particulière à la structuration de leurs données, et particulièrement ceux n'ayant pas encore de données structurées. Les contributeurs ayant déjà un minimum de données structurées n'ont pas observé de différence particulière lorsque de nouvelles données ont été structurées.

La structuration de données est également pertinente pour les contributeurs qui planifient et publient leur programmation en amont, laissant alors une période suffisante aux événements de se laisser découvrir par les publics. Dans un contexte de pandémie puis de post-pandémie, où la reprise des activités du secteur culturel est une affaire de tous, il est plus ardu de quantifier une véritable évolution.

## Perspectives d'amélioration

Pour établir une causalité entre la structuration des données et le nombre de pages vues, deux ajustements s'imposent. D'une part, il est nécessaire d'obtenir un échantillon conséquent de données pour tous les contributeurs, avec pour chacun au moins une saison de données non structurées pour comparaison. D'autre part, le maintien d'un groupe de contrôle, constitué de pages d'événements aux données non structurées, permettrait une comparaison précise. Ces modifications pourraient accroître la validité de l'étude.

# Références

Institut Statistique du Québec. (2021). *Conception d'indicateurs de rayonnement de l'offre culturelle québécoise dans les réseaux numériques.*

<https://statistique.quebec.ca/fr/fichier/conception-indicateurs-rayonnement-offre-culturelle-quebecoise-reseaux-numeriques-faisabilite.pdf>

Ministère de la culture et des communications du Québec. (2021). *Guide de bonnes pratiques : découvrabilité et données en culture.*

<https://cdn-contenu.quebec.ca/cdn-contenu/adm/min/culture-communications/publications-adm/publications-complementaires/GM-Bonnes-pratiques-decouvrabilite.pdf>

Plamondon, J. (1er août 2022). *Découvrabilité : Oui, mais dans quel environnement technologique?*

<https://joseeplamondon.com/decouvrabilite-environnement-technologique/>

Printemps numérique. (28 octobre 2021). *Conférence—Les défis à venir pour la découvrabilité des contenus francophones en ligne.*

<https://www.youtube.com/watch?v=bo3nXHFTAHO>

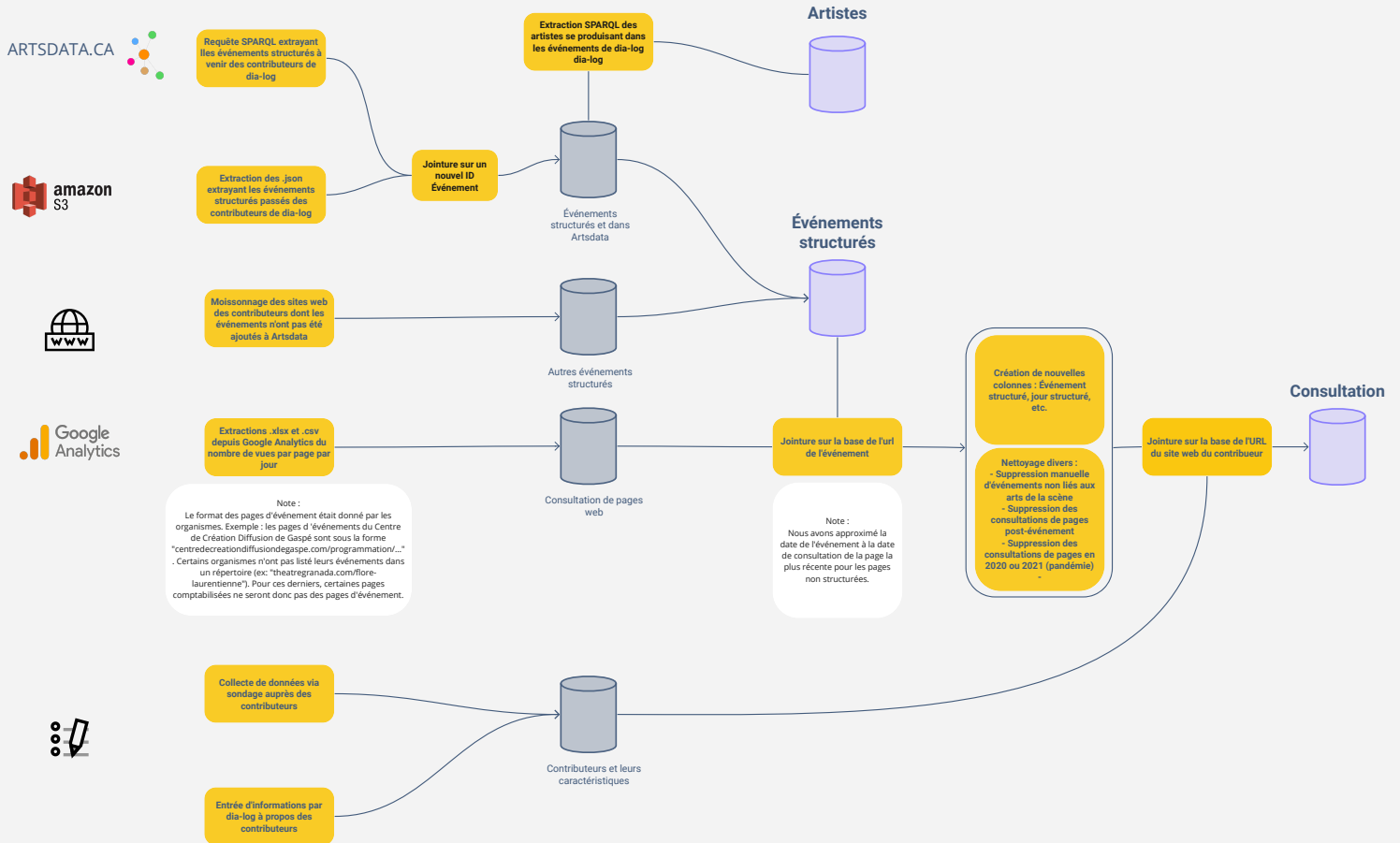
Rioux, M., & Mili, H. & Tremblay, D. (2021). *Mesure de la découvrabilité des produits musicaux et audiovisuels québécois sur les plateformes numériques.*

[https://frq.gouv.qc.ca/app/uploads/2021/10/michele.rioux\\_decouvrabilite-plateformes-numeriques\\_rapport-de-recherche.pdf](https://frq.gouv.qc.ca/app/uploads/2021/10/michele.rioux_decouvrabilite-plateformes-numeriques_rapport-de-recherche.pdf)

Wells, G.-P., & Desjardins, D. (2020). *Pratiques culturelles numériques de promotion, de diffusion et de monétisation du contenu francophone canadien sur Internet. Tendances, obstacles et opportunités.* 26.

# Annexes

# Annexe 1. Schéma de transformation des données



## **Annexe 2.** Limites des données à disposition

**BD Événements :** Certains événements n'avaient pas le même ID Événement alors qu'ils réfèrent au même événement. Ils ont donc été fusionnés sur la base d'un nouvel ID Événement composé de la concaténation de l'URL de la page et de la date de l'événement. Cette technique n'est pas infaillible et il est possible que la base de données d'événements finale ait des doublons.

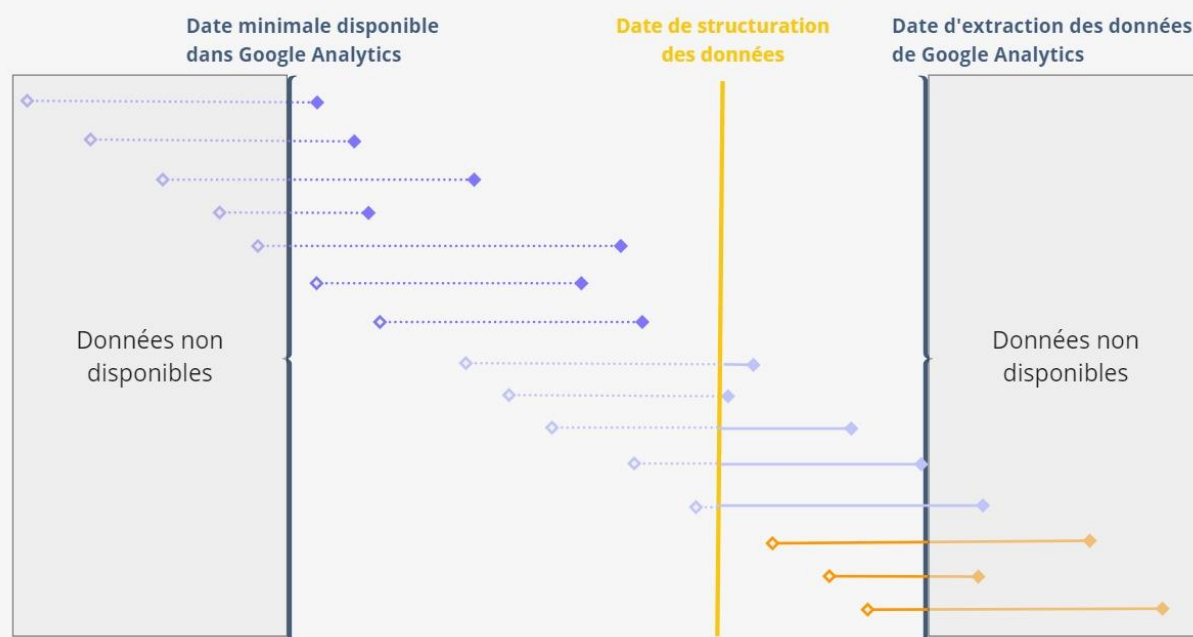
**BD Artistes :** Dans certains cas, la requête d'extraction des informations sur les artistes ne renvoyait pas de résultats malgré l'existence de l'entité et des informations dans le graphe de connaissances (cela représentait 15 % des artistes recensés). Les raisons de ce phénomène sont inconnues. Cela peut avoir une influence sur le potentiel de nombre de liens, qui pourrait être sous-évalué.

### **BD Consultation :**

- Données censurées et ce de manière inégale : Les données de pages vues (issues de Google Analytics) n'étaient pas toujours disponibles pour une saison entière d'un contributeur à l'autre. Cela suppose que :
  - 1) nous comparons des événements ayant parfois lieu à des périodes observant naturellement plus de visibilité que d'autres;
  - 2) les débuts et fins de période de mise en ligne doivent parfois être inférés puisque les événements structurés auront tendance à n'avoir que le début de la fenêtre de mise en ligne, et les événements non structurés n'ont parfois que la fin de leur fenêtre de mise en ligne (voir annexe 3).
- Bruit dans les données : Il n'est pas toujours garanti qu'une page soit bien une page d'événement, et particulièrement une page d'événement en arts de la scène. Bien qu'un effort ait été fait au fur et à mesure de l'étude pour éliminer les erreurs lorsqu'elles étaient identifiées, cela représente un travail important de nettoyage manuel. Ainsi, certaines pages peuvent avoir été comptabilisées dans l'étude sans toutefois être des événements spécifiquement en arts de la scène.

### Annexe 3. Définitions d'un événement structuré, non structuré et semi-structuré

Le rayonnement d'un événement se mesure par l'intérêt qu'il suscite avant d'avoir lieu. En effet, par sa nature éphémère, il diffère d'autres objets culturels et il n'est donc pas pertinent d'évaluer sa découvrabilité une fois celui-ci terminé. Afin d'évaluer la structuration des données, nous observons le nombre de pages vues selon 3 types d'événement.



- ◆—◆ Événements non structurés (passés)
- ◆—◆ Événements partiellement structurés
- ◆—◆ Événements structurés
- ◇ Date de mise en ligne de l'événement
- ◆ Date de l'événement
- Jour structuré
- ..... Jour non structuré

#### Un événement structuré

désigne un événement dont les données sont structurées depuis que celui-ci a été mis en ligne.

#### Un événement non structuré

désigne un événement passé dont les données n'ont pas été structurées avant la date où celui-ci a eu lieu.

#### Un événement semi-structuré

désigne un événement dont les données n'étaient pas structurées à la date où celui-ci a été mis en ligne, mais les ont été avant la date de l'événement.

Puisqu'il existe des événements semi-structurés, nous pouvons observer au niveau du jour si les données d'un événement étaient structurées ou non.

**Un jour structuré** désigne un jour où un événement avait des données structurées.

**Un jour non structuré** désigne un jour ou un événement n'avait pas encore de données structurées.