



La Lettre de l'OCIM

Musées, Patrimoine et Culture scientifiques et techniques

172 | 2017
juillet-août 2017

Connaître ses visiteurs en ligne : quels outils, quelles méthodes ?

Florence Andreacola, Éric SanJuan and Jean Guibal



Electronic version

URL: <http://journals.openedition.org/ocim/1823>

DOI: 10.4000/ocim.1823

ISSN: 2108-646X

Publisher

OCIM

Printed version

Date of publication: 1 July 2017

Number of pages: 5-11

ISSN: 0994-1908

Electronic reference

Florence Andreacola, Éric SanJuan and Jean Guibal, « Connaître ses visiteurs en ligne : quels outils, quelles méthodes ? », *La Lettre de l'OCIM* [Online], 172 | 2017, Online since 01 July 2018, connection on 07 May 2019. URL : <http://journals.openedition.org/ocim/1823> ; DOI : 10.4000/ocim.1823

This text was automatically generated on 7 May 2019.

Tous droits réservés

Connaître ses visiteurs en ligne : quels outils, quelles méthodes ?

Florence Andreacola, Éric SanJuan and Jean Guibal

Lever les yeux est un prototype imaginé au musée Dauphinois comme une expérience entre contemplation et médiation.



© Musée Dauphinois/Denis Vinçon

- 1 Visiter une exposition dans un espace de réalité virtuelle ; arpenter, à l'aide de sa souris, une salle d'exposition reconstituée en profondeur sur son écran d'ordinateur ; explorer une salle par robot interposé ou encore faire l'expérience d'une œuvre grâce à un site Internet complémentaire à l'exposition physique sont autant d'expériences de découverte que proposent certains musées sur Internet. Il est aujourd'hui acquis que la plupart des

musées possèdent un site Internet à partir duquel un internaute peut découvrir la programmation, les collections, les grandes thématiques abordées, certains objets "phares" de la collection, ou encore des descriptifs plus ou moins développés des expositions temporaires ou permanentes. Et dans tous les cas, ces sites Internet sont a minima chargés de délivrer les informations pratiques pour se rendre au musée.

- 2 Les musées peuvent aussi déployer leurs actions et leur offre de médiation sur les réseaux sociaux numériques, sur les plateformes de partage de vidéos ou de sons. Ils peuvent également être présents sur les portails de mise en commun des collections ou de la programmation culturelle d'un territoire à différentes échelles. Il ne fait pas de doute que le musée essaime sur Internet vers de nouvelles formes de visibilité, de transmission et de médiation. Dans ce contexte, comment connaître le public de ces multiples initiatives qui se déploient sur le territoire non cloisonné d'Internet ? Comment mesurer les effets de ces actions disséminées sur des espaces plus ou moins privés du web, et notamment les effets sur la visite même de l'établissement ? Il est certain que toute forme d'enquête par entretien, questionnaire ou observation qui aurait lieu au musée peut difficilement atteindre les publics exclusivement internautes du musée, s'ils existent.
- 3 L'objet de cet article est de présenter les possibilités concrètes que peut mettre en œuvre un musée pour mieux connaître les effets de ses actions en ligne ainsi que les publics qui le fréquentent par Internet. Ces propositions d'outils et de méthodes se basent sur les expérimentations et les résultats d'une recherche universitaire menée en partenariat avec un musée sur ses publics en ligne et *in situ*.
- 4 Nous présenterons, ci-dessous, la nature spécifique des données informatiques qui sont produites lors des usages d'Internet et comment elles nous renseignent sur ses usagers. Dans un deuxième temps, nous détaillerons les formes de dissémination du musée en ligne en décrivant les problématiques d'accès aux données d'usages. Ensuite, nous expliquerons le partenariat mis en place qui a permis d'avancer sur ces problématiques. Enfin, nous aborderons les modalités concrètes d'accès aux données pour présenter une conclusion sur les usages croisés et potentiels de ces données.

Un musée en ligne protéiforme

- 5 Les publics du musée sont potentiellement confrontés à de nombreux avatars du musée numérisé. Alors que la plupart des musées possèdent un site Internet entièrement dédié à l'institution, ils déploient aussi leur communication et leur médiation en ligne sur d'autres sites Internet. Ces sites sont souvent choisis pour leur spécialisation médiatique : jouer des vidéos (YouTube, Vimeo), déposer des enregistrements sonores (Soundcloud), enregistrer un nombre important d'images en haute définition (Flickr, Wikimedia), déposer des fichiers comme des dossiers pédagogiques, des dossiers de presse (Issuu, Calameo). Ils peuvent aussi être choisis dans une logique de multiplication des accès : utiliser un site Internet avec une forte visibilité et un fort taux d'usages peut faire connaître les activités du musée par un public potentiellement plus important. On pense aux plateformes que sont Facebook, le Google Art Project ou encore Wikipédia. Les musées peuvent aussi choisir de mutualiser leurs ressources scientifiques et culturelles par le biais de plateformes publiques généralistes comme Europeana, le Moteur Collections ou encore la base de données Joconde. Ils peuvent aussi privilégier des plateformes publiques spécialisées comme Vidéomuseum pour l'art contemporain ou encore araGo, un portail dédié à la photographie.

- 6 Nous relevons quatre formes de présence du musée sur Internet. Tout d'abord, nous désignons le site Internet du musée comme une entité centrale. À cette entité centrale s'ajoutent les plateformes publiques qui ont plutôt pour mission de contribuer à la valorisation et à la diffusion du patrimoine par le biais de la collection. Ensuite, les musées exploitent des plateformes privées qui sont des outils de publication spécialisés par formats médiatiques (comme You tube pour la vidéo) ou davantage destinées à renforcer les processus relationnels du musée (comme Facebook ou Twitter). Nous faisons l'hypothèse que ces formes de présence de l'institution muséale sur Internet complexifient l'image d'une institution muséale "numérisée" par la variété des portes d'entrée sur son contenu. Par ailleurs, ces formes démultiplient mécaniquement les stratégies éditoriales de l'institution. Enfin, cette dissémination brouille l'accès aux données relatives aux usages des publics et diversifie leur format. Afin de mieux cerner les chemins d'accès et la valeur de ces données d'usage, nous détaillons dans cette partie les caractéristiques spécifiques des formats et modes d'accès aux données d'usage en fonction d'un regroupement en trois points : le site Internet du musée, les plateformes publiques et les plateformes privées.

Le site Internet du musée : une entité centrale ou un double numérique ?

- 7 Le site Internet du musée peut être administré en interne, par des membres du personnel du musée ou de la collectivité de tutelle. Son administration peut aussi être soumise à des formes de délégation ou de sous-traitance. En fonction des formes de délégation, le musée aura plus ou moins la possibilité d'effectuer des modifications sur son site. Il aura aussi un accès aux journaux de connexions. C'est donc en fonction du contrat signé avec un prestataire éventuel que le musée aura accès ou non aux données d'usages de son site. Dans la plupart des cas, la configuration la plus favorable pour obtenir et traiter ces informations est celle qui limite le nombre d'intermédiaires et celle qui établit des liens privilégiés avec le service informatique de la collectivité de tutelle. Car c'est souvent par elle que la gestion de ce type de service s'établit. Les informations relatives aux usages du site web sont précieuses pour connaître les publics du musée. Dans ce contexte, elles deviennent des données informatiques de première main qui témoignent de manière codifiée de la relation qui s'établit entre le musée et ses internautes. Dans le cas d'une situation où le musée, soit directement, soit par le biais de sa collectivité de tutelle, administre son site Internet, ces données sont jugées complètes. C'est-à-dire que l'ensemble des actions réalisées par un internaute sur ce site est enregistré et peut être exploité par le musée dans le respect de règles de confidentialité habituelles.

Capture d'écran sur You Tube d'une vidéo de la chaîne du musée Dauphinois : La Machine à contes est un dispositif interactif multimédia.



La page d'accueil de la base de données photographique araGo (www.photo-arago.fr).



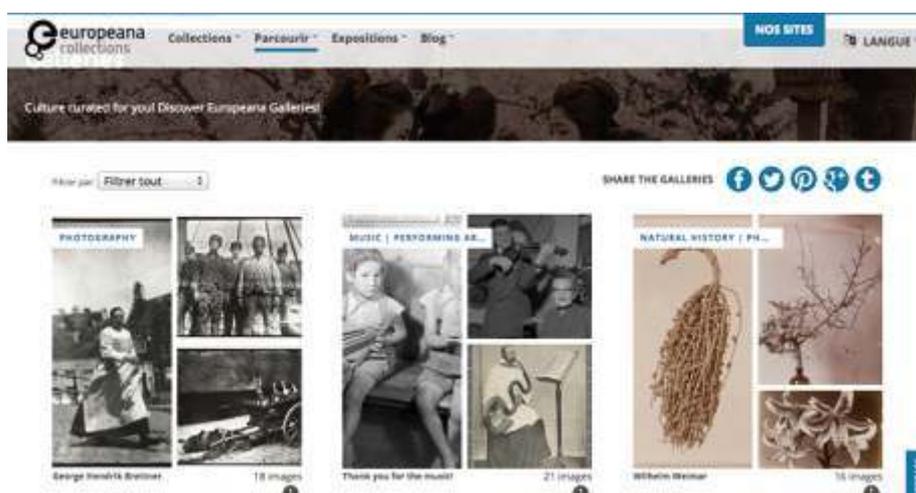
Le site Calameo héberge une ressource de médiation du musée de la Camargue (<http://fr.calameo.com/>).



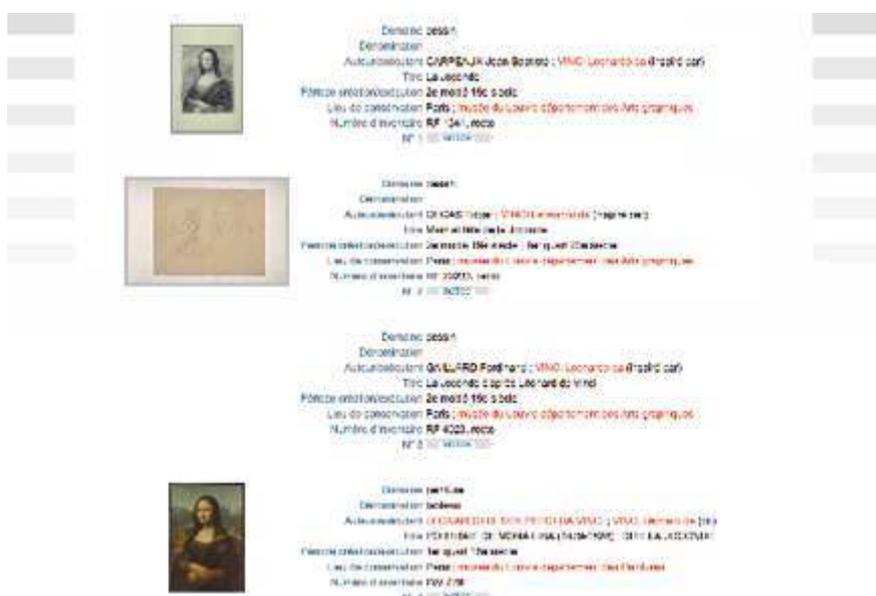
Les plateformes publiques : diffusion et valorisation du patrimoine par la mise en commun

- 8 Alors que le chemin peut être relativement court entre le public et le musée lorsque ceux-ci communiquent via le site Internet du musée, la relation qui s'établit entre un internaute et un ou plusieurs objets de la collection de l'institution devient plus complexe lorsqu'elle passe par une plateforme publique qui œuvre à la diffusion et la valorisation d'un patrimoine culturel. Étant donné que le propriétaire du site Internet a accès aux données d'usage, il est donc libre de les transmettre à ses partenaires et selon le format qu'il souhaite. Ces échanges d'information sont donc inégaux d'une situation de partenariat à l'autre. Par exemple, dans le cas de la plateforme Europeana, les journaux de connexion sont accessibles sur demande et font l'objet de plusieurs recherches scientifiques¹. La base Joconde, quant à elle, diffuse quelques éléments issus des journaux de connexion comme les statistiques relatives à la fréquentation du site, de ses expositions virtuelles et des images vues. Ces informations sont partielles, mais accessibles à tous depuis le site de la base Joconde². Il est toutefois assez rare d'avoir accès à des données relatives aux traitements des journaux de connexions sur les plateformes culturelles.

La base de données Europeana (<https://portal.europeana.eu/>)



La base de données Joconde (www.culture.gouv.fr/public/mistral/joconde_fr)



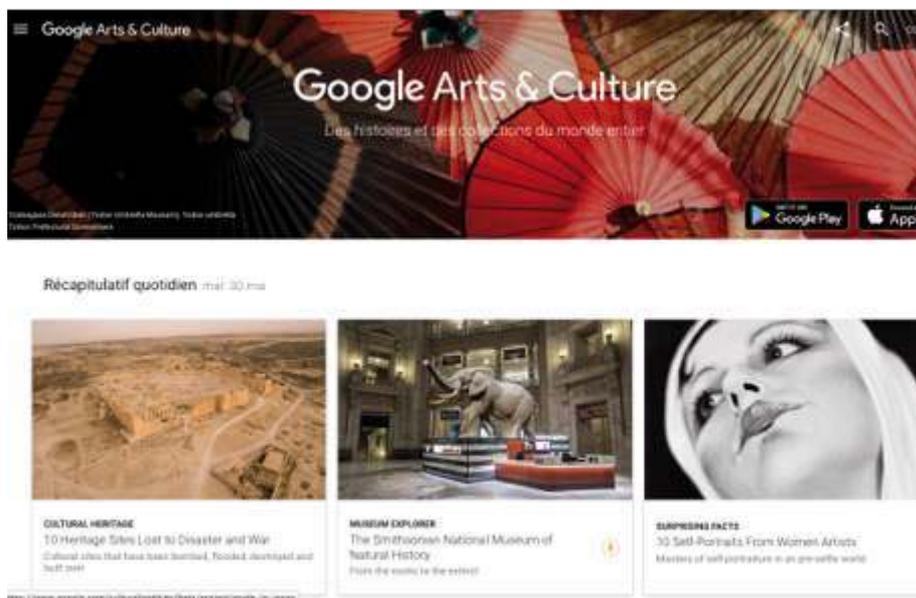
Les plateformes privées : éditorialisation par média et spécialisation de la fonction relationnelle du musée

- 9 Une vidéo qui présente la bande-annonce d'une exposition temporaire sur YouTube ; l'interview d'un conservateur présentant un objet de la collection sur une chaîne de télévision locale et diffusée en *replay* sur le site de la chaîne ; la collection en haute définition sur Google Art Project ; les documents pdf du dossier de presse ou du dossier pédagogique lisible sur une plateforme de lecture comme Issuu sont souvent relayés sur une page Facebook ou un compte Twitter de l'institution. Ces deux plateformes "sociales" sont aussi le lieu des annonces événementielles du musée. Ces modes de stockage, de gestion et de diffusion de contenus culturels sur Internet sont autant d'opportunités saisies par une institution publique qui, en échange d'un service de publication gratuit,

cède à des plateformes privées des contenus scientifiques et culturels ainsi que des données relatives à leurs usages.

- 10 Néanmoins, ces plateformes livrent des indications comme le nombre de fois qu'une vidéo a été visionnée, des bilans de "like" ou de formes "d'engagement". Ces plateformes, qui ont la caractéristique de normaliser les formats d'informations qu'elles acceptent de diffuser aux usagers, choisissent aussi les informations qu'elles acceptent de diffuser parmi l'ensemble des données relatives aux usages des internautes de leurs services. Ces données peuvent être particulièrement précises et riches lorsque les internautes sont soumis à une identification comme la création d'un compte avec adresse mél, mot de passe et quelques données personnelles. C'est par exemple le cas de Facebook ou de Twitter. Sur Facebook, on assiste à la mise en place de service de mesure d'audience ou "d'engagement" à disposition des administrateurs de pages. Un groupe de chercheurs néerlandais (INTK) a d'ailleurs mis en place, sur la base de données collectées via les comptes des institutions, une publication des résultats d'adhésion (à partir du nombre de suiveurs – traduction de *followers*, devenu un anglicisme – ou d'abonnés) aux réseaux sociaux par les internautes. Il s'agit de Museum Analytics³. Pour chaque musée inscrit à ce projet, des comparaisons sont établies entre les chiffres de fréquentation du site physique et les chiffres du nombre d'abonnés aux comptes Twitter et aux pages Facebook des institutions participantes. Ces données peuvent paraître très utiles au *community manager* d'une institution muséale qui voit en temps réel certains effets des actions de publication en ligne. Sur Facebook, depuis un back-office dédié aux statistiques d'usages, il peut en extraire des tableaux de données pour défendre auprès de sa hiérarchie l'efficacité de certaines actions, leur visibilité, le taux "d'engagement" qu'elles suscitent. Dans ce contexte, il est important de connaître les enjeux relatifs à leur production pour les interpréter à leur juste valeur. Cette prudence s'impose, car les objectifs de ces plateformes privées sont essentiellement commerciaux. Elles délivrent donc des informations sur les usages en fonction de stratégies éditoriales et commerciales : vendre plus d'espace publicitaire, encourager les usages pour engranger des données... Ces informations sont donc partiales. Par ailleurs, les informations qui existent en ligne au sujet d'un objet de la collection ou des actions du musée ne sont pas exclusivement produites par le musée. Le cas de figure où des internautes racontent leur visite sur leur blog de voyage ou encore affichent une photographie prise dans l'exposition sur leur page Facebook, pour ne citer que quelques exemples, sont difficiles à prendre en compte. Ces contenus, soumis à des systèmes complexes de confidentialité, sont extrêmement difficiles à collecter de manière exhaustive pour le musée. S'éloigne donc la possibilité d'atteindre les usages de ces contenus produits par des internautes pour d'autres internautes au sujet du musée dans une approche rigoureuse.

La page d'accueil du Google Cultural Institute (www.google.com/culturalinstitute)



Une démarche d'évaluation globale

- 11 Dans ce contexte, comment un musée peut-il construire ou s'approprier les outils nécessaires pour mieux connaître ses visiteurs-internautes ? Peut-il s'appuyer sur les données transmises par les plateformes privées d'édition de contenu ? Peut-il faire bon usage des statistiques parcellaires et publiques des plateformes privées ?
- 12 Dans le cadre d'une recherche doctorale, nous avons décidé d'explorer ces questions en partenariat avec un musée de société, le musée Dauphinois à Grenoble. Cet établissement est doté d'un site Internet géré par les membres de son équipe qui animent en outre une page Facebook, un compte Twitter et diffusent des contenus sur des plateformes publiques et privées. Le musée a, par le biais du service informatique du Département de l'Isère (qui assure sa tutelle administrative), un accès direct aux données d'usages de son site Internet. Jusqu'alors, le musée se tenait informé des activités d'usage par Google Analytics. Dans le cadre de ce partenariat, nous avons exploré, d'autres pistes de recherches sur ces données tout en menant une réflexion critique sur les outils. Cette partie est consacrée à la présentation de la forme du partenariat, à l'approche mise en œuvre pour étudier le journal de connexion du site Internet du musée et à la conception d'un questionnaire en ligne.

Récit d'un partenariat

- 13 Alors que tous les établissements culturels ont mis en chantier des études en vue de mieux connaître leurs publics, beaucoup ont relevé le risque d'une évolution des politiques culturelles vers le marketing, c'est-à-dire l'élaboration d'un projet établi d'après la demande des publics et non plus de l'offre. Ce débat doit être entretenu et les responsables d'institutions culturelles sont tenus de rester vigilants. Mais la connaissance

des publics, comme l'étude de leurs pratiques et de leurs usages du musée reste une nécessité absolue, même si l'on n'envisage pas une finalité commerciale. On ne peut plus gouverner un établissement culturel, on ne peut plus prétendre travailler à l'élargissement de ses publics si l'on ne dispose pas d'outils de mesure adéquats pour des pratiques culturelles qui existent aussi en ligne. Le Département de l'Isère dont relève le musée Dauphinois et qui compte un réseau de dix musées, a conduit ces derniers mois une grande étude (confiée à l'Observatoire des politiques culturelles, qui s'est assuré la collaboration de plusieurs chercheurs, dont les résultats seront publiés en 2017), tant quantitative que qualitative, sur la fréquentation de ses musées sur la durée d'une année complète. Dans le même temps, les pratiques de muséographie "participative" mises en œuvre dans ce musée, comme ses expériences dans le domaine de la culture numérique, attiraient une chercheuse en muséologie, qui proposait d'explorer la rencontre des publics par le biais des technologies numériques. Sous sa direction, une première observation était tentée lors d'une expérience originale de médiation au moyen de ces technologies (Muséomix) qui concluait d'ailleurs à relativiser l'intérêt pour le public, dans le musée, de l'usage de ces technologies.

- 14 Dès lors, une étude plus précise et plus ambitieuse était programmée, pour laquelle le musée Dauphinois s'associait à un laboratoire de l'université d'Avignon et des Pays de Vaucluse (la Fédération de Recherche CNRS Agorantic) et permettait aux chercheurs d'avoir un accès complet aux données des usages du site Internet du musée Dauphinois. Au-delà de l'observation de la fréquentation des sites Internet des musées, il s'agissait de mieux comprendre et analyser les usages des outils numériques et de tenter d'en livrer une typologie. Autant de résultats de nature à influencer sur la conduite du musée, dans ses expositions comme dans ses pratiques d'Internet.

Étudier le journal de connexion

- 15 Google propose un service d'étude des usages d'un site web qui se nomme Google Analytics (GA). Ce service permet d'analyser la fréquentation d'un site Internet gratuitement. La version gratuite de GA se base sur la captation, à l'aide d'un code ajouté par l'administrateur du site à chacune de ses pages (*trackeur*), des usages du site et de leur transmission aux serveurs de GA. Si l'utilisateur a installé des outils de non-traçabilité qui empêchent ces enregistrements, GA perd ces données. Par ailleurs, afin de réduire le temps de réponse des calculateurs, GA base ses résultats sur des échantillons de données et non pas sur la totalité des données en sa possession. Les données que transmet GA gratuitement sur les usages d'un site Internet sont par ailleurs censées favoriser l'achat d'espaces publicitaires dans le but d'accroître la "performance" du site Internet. Soumises à ces stratégies commerciales, ces données statistiques transmises à l'administrateur du site web sont à lire dans ce contexte, en tenant compte de ces enjeux commerciaux. Par ailleurs, il est sans doute nécessaire pour le musée qui installe de sa propre initiative ces *trackeurs*, de s'interroger sur la valeur des données qu'il donne de son plein gré à Google. Dans le cas du recours à ce service, le musée livre à Google des données au sujet de ses publics et de la relation qui s'établit avec eux par le biais de son site Internet et de son offre culturelle. Grâce à ces données et leur analyse, Google a compris, par exemple, que les internautes se rendent sur le site Internet d'un musée en majorité pour connaître les informations pratiques : localisation géographique, horaire d'ouverture, tarif et numéro de téléphone. Fort de cette connaissance, Google affiche désormais ces informations

directement sur son moteur de recherche. Dans ce contexte, le site Internet du musée risque de voir baisser sa fréquentation, les internautes s'arrêtant avant sa porte numérique.

- 16 Dans notre recherche, nous avons pris le parti d'analyser les usages des internautes à partir de la source que constitue le journal de connexion transmis par l'hébergeur. Les données brutes sont difficilement exploitables en l'état. En effet, le site analysé reçoit en moyenne 5 000 visiteurs uniques par mois, ce qui donne un nombre considérable de navigations, de chemins d'entrée, de sortie pour ne citer que quelques aspects analysables. Pour pouvoir les exploiter statistiquement, il est nécessaire de travailler avec un corpus important de données (de l'ordre de plusieurs mois dans le cas du musée Dauphinois). Un premier outil de traitement statistique peut être utilisé comme l'outil Awstat. Cet outil permet d'analyser les données en local (sur sa propre machine) et en "hors connexion" ce qui permet de garantir la protection des données. Cet outil est un logiciel libre et gratuit. Il permet de mettre en évidence la fréquentation du site web, les mots clés utilisés pour y accéder, le temps passé par page, les pages les plus fréquentées, les origines géographiques des internautes, les systèmes d'exploitation, les navigateurs, les éléments téléchargés, les pages sur lesquelles la visite s'est arrêtée, les pages sur lesquelles la visite a démarré... Cette liste non exhaustive illustre les possibilités d'analyse de ce type d'outils et permet de mieux connaître : les sujets d'intérêt, la géographie, les profils technologiques de ses publics en ligne. Néanmoins, avoir à disposition ce type d'information, presque en temps réel, renforce le risque de se perdre dans les données. Comment juger celles qui sont pertinentes par rapport à la stratégie du musée ? Lorsque l'on utilise un outil de traitement statistique de ces données issues du web, qu'elles soient fournies par GA, le journal de connexion traité par Awstat ou Facebook, nous n'évudons pas la question fondamentale de la définition d'indicateurs pertinents. Le musée est alors face à la responsabilité de les définir, éventuellement en lien avec des chercheurs.

Pendant l'expérience Museomix, le prototype de Tout schuss fait voyager le visiteur du musée Dauphinois dans le temps et sur piste en immersion.



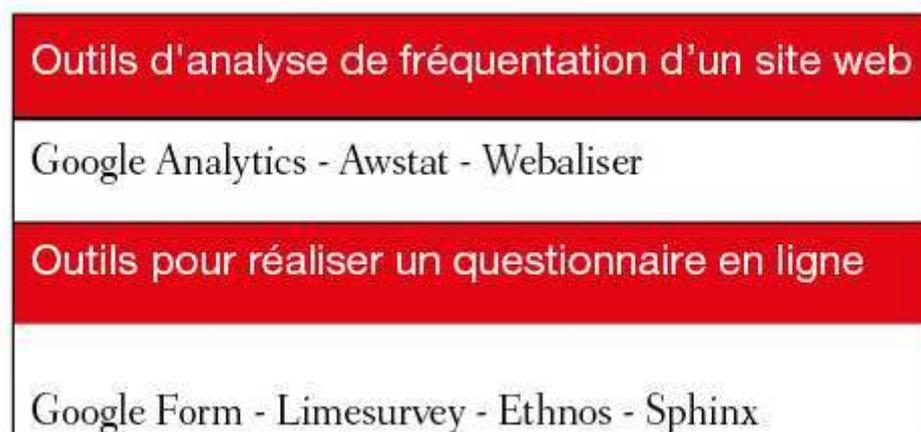
© Musée Dauphinois/Denis Vinçon

Concevoir et administrer un questionnaire en ligne

- 17 Une autre possibilité pour mieux connaître les internautes-visiteurs du musée et les effets de ses actions sur Internet est de réaliser une enquête par questionnaire en ligne. Bien entendu, il est possible d'atteindre certains résultats par des outils d'enquête habituels comme l'entretien ou le questionnaire en fin de visite auprès de visiteurs bien réels venus au musée. Le questionnaire administré en ligne ne remet pas en question la validité et l'efficacité de ces modes d'enquête. Par contre, l'enquête en ligne permet d'atteindre et d'évaluer une forme spécifique de participation en ligne par la prise en compte du temps passé à répondre : ce temps témoigne, selon nous, de l'engagement du répondant à entrer en relation avec le musée de manière explicite. Ce format d'enquête permet aussi un traitement immédiat des résultats, sous la forme de statistiques descriptives. Par ailleurs, ce format offre au musée la possibilité d'effectuer une administration régulière (annuelle ou bisannuelle) du questionnaire auprès de son public-internaute. Il est important de tenir compte du faible taux de réponse pour ce type d'outil : de l'ordre de 20 %. Il nécessite donc une expertise dans la construction du questionnaire, sa diffusion et une analyse fouillée des résultats. Là encore, il existe déjà plusieurs types d'outils disponibles en ligne et qui sont quelquefois gratuits : Google Form ou encore les logiciels professionnels Ethnos ou Sphinx proposent des solutions pour mener des enquêtes en ligne. Dans le cadre de notre recherche, nous avons utilisé le logiciel libre et gratuit Limesurvey. Cette configuration permet un hébergement du questionnaire sur un serveur indépendant et garantissant la confidentialité des données. Nous avons remarqué un plus

haut taux de réponse lorsque le questionnaire était administré aux abonnés de la lettre d'information du musée. En effet, l'administration à partir du site Internet avec une invitation présente sur la page Facebook du musée a connu un très faible taux de questionnaires remplis. Ce type de procédure, d'une apparente rapidité et facilité, nécessite certains savoir-faire dans la construction d'un questionnaire, des tests techniques pour sa bonne diffusion, d'un suivi pour les demandes spécifiques (ce type de questionnaire engage dans une relation) ainsi que dans l'analyse en profondeur des résultats.

Quelques outils d'analyse des logs et de conception et d'administration de questionnaire en ligne.



Capture d'écran d'une vidéo sur l'exposition Confidences d'outre-tombe de la chaîne You Tube du musée Dauphinois



Conclusion

- 18 Cette recherche nous a amenés à expérimenter des méthodes et à interroger des outils d'analyse dont certains sont d'apparence facile à mettre en œuvre et gratuits. Les méthodes et les outils qui sont décrits et analysés ont notamment pour objectif d'atteindre des activités de réception et d'usage de la culture en dehors de l'espace muséal. On s'aperçoit que les contraintes et leviers sont à la fois pratiques, techniques et méthodologiques pour obtenir un panorama riche, complet et fiable de ces usages en ligne et s'éloigner d'une évaluation basée uniquement sur le nombre de vue ou de *like*.
- 19 Cette approche ne doit pas faire oublier que le musée est un outil culturel indispensable dans nos sociétés contemporaines. Cette nécessité d'un lieu de préservation et de rencontre avec le document original – qu'il s'agisse d'une œuvre d'art ou d'un simple témoignage d'une forme de culture – l'est chaque jour plus encore avec la progression des technologies de la représentation numérique. Autrement dit, aucune de ces technologies, si efficace et si valorisante soit-elle, ne remplacera la rencontre avec l'original et l'émotion qu'elle procure à chacun tel qu'il est possible de le vivre au musée.
- 20 Il n'en reste pas moins vrai que la médiation que permet Internet, pour toutes les formes de présentation du patrimoine, est un atout majeur pour les musées et les services de conservation en général. Aussi est-il indispensable que ces établissements culturels soient présents sur le réseau mondial, et qu'ils s'y présentent de la meilleure façon. Ce qui implique que l'on doit tenter de comprendre les pratiques des internautes, qui sont souvent bien éloignées des intentions des concepteurs des sites. Le musée est donc amené à considérer et connaître ce public particulier, qu'il confirme ou non son intérêt par la visite des expositions *in situ*.
- 21 Cette contribution n'a d'autre ambition que d'attirer l'attention des responsables de musées sur les usages d'Internet et plus largement des outils numériques (dans les salles du musée comme chez soi) et de proposer une méthodologie⁴ pour en conduire une précise observation.
- 22 *Le texte suivant figure en encadré dans l'édition papier :*

Le format des données du web

- 23 Le musée qui souhaite connaître au mieux ses visiteurs et la façon dont ils reçoivent ses projets culturels, peut être tenté par la possibilité de collecter et d'analyser l'ensemble des informations véhiculées sur Internet à son propos. Mais ce travail se heurte alors à la question de l'accès, de la propriété et de la complétude des données. Le musée sera ensuite confronté à la question des outils méthodologiques et critiques nécessaires à leur exploitation. Pour parvenir à dépasser ces verrous, il nous semble utile d'envisager les données relatives à l'activité culturelle en ligne des internautes d'une manière non circonscrite à ses aspects immédiatement visibles à l'écran que sont un commentaire écrit au sujet d'une photographie sur Facebook, la publication de photographies de sa visite sur une plateforme telle que Flickr, ou encore le récit d'une visite d'exposition sur un blog personnel. Dans cette logique, nous présentons dans cet encadré les formats des données de l'Internet et ce qui les régit.

Protocole d'Internet

- 24 L'Internet se base sur un protocole d'échange développé par la DARPA (Defense Advanced Research Projects Agency) qui a pour objectif de développer un moyen fiable de transmettre des messages sur un réseau informatique. Pour cette raison, l'agence met en place des protocoles d'échanges (TCP, Transmission Control Protocol et IP, Internet protocol) qui ont la particularité de découper les messages en plusieurs paquets. Ces derniers sont alors envoyés de façon démultipliée afin de garantir la transmission. Chaque paquet est identifié par son destinataire et son expéditeur. S'ajoutent à ces deux premières couches d'informations nécessaires aux échanges, le protocole FTP qui permet l'envoi de fichiers, et les protocoles HTTP/HTML qui permettent de traiter la visualisation de médias et l'association de messages issus de sources différentes sur une même page. Les normes de ces protocoles sont fondées sur un consensus et donc un travail d'élaboration collectif stable. Nous souhaitons mettre en évidence que chaque paquet d'information échangé sur le réseau Internet est accompagné d'une série d'écrits contextuels nécessaires à l'efficacité de la transmission et à son format (image, texte, vidéo, fichier). Ces informations contextuelles sont, à l'origine, conservées pour résoudre d'éventuels dysfonctionnements dans la transmission des messages. Aujourd'hui, la loi française impose la conservation de ces données par l'hébergeur du site Internet et son propriétaire pendant une durée d'un an (décret n°2011-219 du 25 février 2011). Ce sont ces données qui nous intéressent particulièrement car elles contiennent à la fois le message et des données contextuelles, rassemblées dans un journal de connexion (en anglais : logs). Elles sont riches car elles peuvent nous renseigner sur les usages que l'internaute aura effectués d'un site Internet. Il s'agit d'une source, qui, lorsqu'elle est prélevée à son origine (sur le serveur d'hébergement), nous garantit sa complétude.

Journal de connexion

- 25 La multiplicité des caractéristiques des données enregistrées sur les activités d'un internaute est d'abord régie par les protocoles d'échange d'Internet qui viennent d'être décrits. Le format et la nature des informations récoltées et enregistrées sur les serveurs hébergeant les sites Internet sont dépendants de ces protocoles et prennent la forme d'un journal de connexion. Les données contenues dans les journaux de connexion des sites Internet présentant des informations relatives au musée, produites ou non par lui, sont donc des sources qui nous renseignent sur les usages des publics du musée présents sur le réseau. Ce journal organise les informations à partir de l'adresse IP (Internet Protocol, soit l'adresse de la machine). Cette adresse nous renseigne sur la localisation géographique (souvent très précise) de l'utilisateur. Reliées à cette adresse IP, les informations relatives au type de moteur de recherche, à la page consultée, au navigateur utilisé et au système d'exploitation sont enregistrées. Chaque action (un clic, une écriture, un téléchargement) est collationnée selon un principe chronologique : une action est reliée à une date, une heure, une minute et une seconde. Dans le cas de figure le plus courant, les logs sont, selon le contrat signé avec l'hébergeur, la propriété du fournisseur du service, en l'occurrence, le musée pour son site Internet.

NOTES

1. <http://labs.europeana.eu/page/logs>
 2. <http://www.culture.gouv.fr/documentation/joconde/fr/partenaires/AIDEMUSEES/statistiques.html>
 3. <http://www.intk.com/en/ideas/museum-analytics>
 4. Pour explorer plus en détail la méthodologie utilisée : Andreacola, F., SanJuan, É. et Poli, M.-S. Méthodologie d'analyse de la participation informatique de l'utilisateur d'un musée, *Revue canadienne des sciences de l'information et de bibliothéconomie*, n°39 (3), 2015.
-

INDEX

Mots-clés: Internet, musée, visite

AUTHORS

FLORENCE ANDREACOLA

docteure en Sciences de l'Information et de la Communication et enseignant-chercheure à l'université de Grenoble-Alpes, équipe Culture et Communication du Centre Norbert Elias (UMR CNRS 8562)
florence.andreacola@univ-grenoble-alpes.fr

ÉRIC SANJUAN

maître de conférences en informatique à l'université d'Avignon et des Pays de Vaucluse, membre du Laboratoire informatique d'Avignon (EA 4128)
eric.sanjuan@univ-avignon.fr

JEAN GUIBAL

conservateur en chef du patrimoine, ancien directeur du musée Dauphinois