

## Musées et Contextualisation Culturelle (MC<sup>2</sup>)

Axe 1 : Culture et numérique

Axe 2 : Réseaux sociaux, structures, contenus et usages

---

### 1) Éléments de présentation

*Laboratoire d'Informatique d'Avignon (LIA - EA 4128)*

**Porteur : Eric SanJuan**

**Collaborateurs :**

**Jean Valère Cossu (doctorant), Pierre Jourlin (MCF), Vincent Labatut (MCF), Juan Manuel Torres Moreno (MCF HDR)**

Ce projet s'appuie sur l'expérience du LIA dans les domaines du Traitement Automatique de la Langue Naturelle (TALN) y compris la Recherche d'information et des réseaux. La présente proposition est dans la continuité de deux projets impliquant le Laboratoire informatique d'Avignon (LIA) liés à l'analyse et à la modélisation du Web, avec des perspectives applicatives d'extraction d'information et de connaissance, de veille et de prédiction des phénomènes sur l'Internet.

- **CAAS** : projet ANR CONTINT, qui a débuté fin 2011 mené par l'IRIT – Toulouse nous a permis de développer les outils de recherche d'information ciblée et contextuelles que nous comptons coupler à une exploration systématique de domaines du Web culturel.
- **ImagiWeb** ; projet ANR CONTINT, qui a débuté en juin 2012 nous permet de développer les outils de détection d'opinion et d'identification de e-réputation sur le Web.

*Centre Norbert Elias (UMR CNRS 8562) ; Équipe Culture et Communication de l'université d'Avignon et des Pays de Vaucluse (ECC)*

**Responsables pour le projet: Marie Sylvie Poli**

**Collaboratrice : Florence Andreacola (Doctorante)**

L'ECC définit pour ce projet la problématique, le terrain d'expérimentation et la méthodologie dans les domaines de la muséologie avec un renforcement de la recherche sur la culture numérique en lien étroit avec l'Agorantic. Ceci dans le cadre contractuel déjà établi avec de nombreux partenaires musées en France et à l'étranger ainsi qu'avec le Ministère de la Culture et de la Communication. En particulier l'UAPV et l'ECC sont membres fondateurs du programme de doctorat international conjoint « Muséologie, Médiation, Patrimoine » (Université du Québec à Montréal – UAPV- École du Louvre).

---

### 2) Descriptif du projet

Le Web sociétal ou Web 2.0 s'est centralisé autour de quelques acteurs privés majeurs dont les bases de données propriétaires permettent d'identifier des segments de réseaux d'individus. Parallèlement à ces réseaux commerciaux s'est affirmé le projet encyclopédique collaboratif Wikipedia fondé sur l'anonymat des contributeurs et l'objectif de mutualisation de savoirs clairement documentés et référencés. De la même manière que le projet Wikipedia a imposé au cœur du Web 2.0 la dimension du savoir collectif, documenté, et ouvert, notre projet est d'insérer la préoccupation des patrimoines culturels matériels et immatériels. Ceci en opposition à la tendance du Web 2.0 de par son modèle économique actuel en niches à se segmenter en petits mondes uniformes tel que le récent projet Google Cultural Institute. Le principe du savoir universel du projet encyclopédique Wikipedia ne permet pas non plus d'aborder cette complexité, y compris dans sa version multilingue actuelle qui tolère des articles différents selon la langue. De fait si le projet encyclopédique continue au fil des siècles à montrer toute sa pertinence comme vecteur du savoir, les musées ainsi que l'ensemble du champ muséal qui ont connu parallèlement autant de révolutions sont aujourd'hui en mesure de devenir de puissants acteurs autonomes du Web 2.0 et d'influer sur celui-ci de part :

- la relation à leurs visiteurs dont on sait qu'ils sont sensibles aux mises en scène des expositions et aux types de médiation de la connaissance. Des visiteurs qui leur consacrent un temps non réductible à un vote électronique ou à une réception passive.
- leurs collections et patrimoines collectifs désormais indexés, référencés et informatisés bien qu'ils ne puissent être que très partiellement virtualisés du fait de la complexité de leurs objets et des contextes culturels associés.

- une présence trans-générationnelle qui contraste avec la succession de réseaux sociaux pour le moment circonscrits à des phénomènes générationnels.
- les réseaux professionnels qui donnent une identité scientifique, symbolique et communicationnelle spécifique à chaque musée et à ses expositions alors que le public, lui, est anonyme par principe. Une inversion de paradigme vis à vis du Web 2.0 où l'utilisateur est ciblé et tracé dans ses moindres actions alors que les responsabilités des principaux sites internet communautaires sont diffuses.
- l'avènement et la reconnaissance de nouveaux acteurs de la culture numérique (Muséomix par exemple) qui défendent le numérique comme nouvelle forme légitime de partage, de médiation et de communication entre le musée et ses publics, réels et potentiels.

**Il s'agit de trouver le dispositif informatique permettant au monde du musée d'accroître sa présence directe sur le Web 2.0 en réduisant autant que possible les intermédiaires. Si dans le cas du projet encyclopédique le simple « wiki », c'est à dire la rédaction collaborative à grande échelle, s'est avéré être le bon dispositif, nous proposons d'appliquer au domaine culturel les méthodes de contextualisation de contenus explorées en Recherche d'Information.**

La contextualisation de messages courts émis sur les sites participatifs, de questions posées par les usagers, de réponses à une question ouverte dans le cadre d'une enquête par questionnaire, ou tout simplement d'un commentaire dans un « livre d'or », consiste à générer automatiquement une note explicative du contexte implicite à ces écrits. Cette contextualisation peut se faire à partir de multiples sources, ce qui permettra ici de définir pour chaque musée un moteur de contextualisation des tendances sur les réseaux sociaux du point de vue de son projet culturel sociétal et patrimonial ainsi que de ses modes de communication et de sa politique d'évaluation de sa programmation.

**Dans un premier temps ce moteur permettra en particulier un échange direct avec les publics via tout type de terminal mobile connecté qui pourra confronter un texte, un message, une réflexion au contexte d'un musée particulier. Si les collections et catalogues d'expositions des musées sont indexées et informatisés, leur interrogation efficace relève d'experts. Le dispositif proposé permettra une interrogation implicite de ces ressources guidée par un contexte d'exposition, qui soit à la fois transparent (sans nécessité de mots clefs précis) et lisible (le texte généré est explicite et complet) pour le visiteur.**

A plus long terme on espère assister :

- une veille culturelle des réseaux sociaux pour les professionnels du musée. Comme il est possible d'évaluer automatiquement la pertinence du contenu des textes générés vis à vis des sources, il sera possible d'insérer la contextualisation dans des systèmes de veille du Web 2.0 auto-adaptatifs par musée. Le professionnel recevra de courtes notes textuelles à fort contenu informatif susceptibles de révéler un contexte culturel implicite à un ensemble d'échanges ou une tendance sur le Web, concernant un musée en général ou une exposition en particulier. La génération de notes textuelles argumentées permettant de révéler des liens complexes allant très au delà de la visualisation de tendances utilisée en marketing.
- une forte présence sur les réseaux sociaux par la libération par le professionnel sur le site du musée des notes automatiquement générées qui lui paraissent informatives. Une fois libérés ces courts résumés générés par extraction et assemblages de données ouvertes mis à disposition par le musée sont propices à être très rapidement indexés par les moteurs de recherche, les sites des musées étant fortement référencés en tant que supports de communication d'institutions culturelles à dimension territoriale. Le contenu textuel indexé, s'il correspond à des textes de qualité, sera propice à se diffuser rapidement dans les différents mécanismes de recommandation du Web 2.0 fondés principalement sur le contenu texte de la requête utilisateur. C'est ce même type de propagation qui a rendu le Wikipedia incontournable aux acteurs du Web. Il s'agira ici de saisir l'opportunité renouvelée pour ces approches de recommandation centrées sur le contenu textuel de la requête depuis le succès de nouveaux moteurs de recherche exclusivement centrés contenu.

### 3) Objectifs et résultats attendus

Notre projet est celui d'esquisser une intelligence computationnelle permettant aux professionnels des musées d'étendre leur action au Web 2.0 dans le cadre de leurs métiers actuels et de leur expertise. L'enjeu consiste à ce que le Musée maîtrise directement sa présence sur le Web sociétal sans nécessiter d'intermédiaires.

En prenant appui sur les données procurées par le Musée Dauphinois dans le cadre contractuel déjà établi avec Agorantic, nous allons étudier la faisabilité de :

- délimiter des sous espace représentatifs de domaines culturels à partir des différents genres de musées et de manière incrémentale avec les professionnels,
- observer automatiquement la sphère d'influence du musée sur Internet suivant une temporalité permettant de restituer la dynamique de son évolution,
- synthétiser ces observations par la génération automatique de notes argumentées et illustrées selon les ressources des musées,
- s'adresser au visiteur pour des interrogations individualisées à sa demande, et au professionnel pour des observations générales,
- observer la réception de ces notes aussi bien par le visiteur que par un expert pour décider de leur diffusion sur ces mêmes réseaux,
- suivre la propagation des notes émises comme autant de sondes, relever leur retour sur les sites des musées au travers des interrogations des visiteurs,
- procéder à un archivage temporel de l'ensemble pour être conservé au musée et être recoupé avec les enquêtes menées auprès des publics (approche anthropologique qualitative et approche quantitative ainsi que modalités d'enquêtes numériques).

### 4) Caractère innovant de ce projet

La centralisation de très larges données individualisées s'est avérée être un atout important en marketing et analyse de l'opinion. La publicité ciblée qu'ils permettent est devenue l'une des principales sources de financements du Web 2.0 tandis que les plateformes de micro blogging telles que Twitter cherchent à se financer sur la revente de ces données. Face à ce double phénomène les institutions culturelles dont les Musées se positionnent de manière ambivalente en y percevant :

- une opportunité d'intéresser de nouveaux publics, de renforcer les liens avec le public existant et d'innover en en muséographie participative et en recherche sur les modes émergents de médiation muséale.
- un risque de basculer dans une logique commerciale de type maximisation d'audience à moindre coûts ou de perdre son identité par dilution dans dispositifs anonymes de type Wikipedia.

**L'expérimentation proposée ici se déroulera en étroite collaboration avec les professionnels du Musée qui pourront suivre ainsi les tendances de l'ensemble du Web 2.0 concernant leurs fonctions au musée au travers d'une interface d'écriture de texte unifiée. Il pourront alors choisir de publier ou pas ce contenu sur le Web et d'en suivre les retours ou l'éventuelle propagation.**

Nous avons identifié cinq types de verrous :

1. De par ses dimensions, le Web ne permet pas une observation indépendante continue et globale. Il n'est pas possible de délimiter un domaine du web en se limitant aux adresses IP ou même aux noms des domaines qui le composent. Leur variabilité et mobilité sont trop importantes. Il n'est donc possible de spécifier les domaines thématiques du Web que de manière dynamique au travers de points d'entrée initiaux, de ressources linguistiques (entités nommées, termes et expressions multi-mots, questions posées) et de critères de style (registres de discours, niveaux de langage et éléments contextuels).
2. Alors que les travaux des chercheurs en muséologie démontrent que le champ muséal est désormais en permanence sollicité et interpellé par des internautes et des pro/ amateurs actifs dans les mondes du numérique et de la culture, les professionnels des musées sont coupés de ces « travailleurs du savoir » qui s'activent sur les réseaux sociaux et les plateformes plus ou moins spécialisés. Il est maintenant opportun de donner à ces réseaux d'acteurs la possibilité de collaborer sans passer par les

- sociétés numériques qui dominent le marché.
3. Une partie croissante du Web est cachée. D'abord les principaux sites internet communautaires tel que FaceBook sont des réseaux privés virtuels cryptés qui utilisent l'infrastructure internet mais dont les entités et les liens n'existent que dans leurs bases de données propriétaires. Ensuite les applications pour terminaux mobiles tel que Android proposées par ces mêmes acteurs peuvent ne pas communiquer pas avec les sites tiers (par exemple le Guardian a communiqué qu'il ne pouvait plus identifier l'origine d'un quart de ses visiteurs). Enfin, le réseau Thor qui permet la navigation sur Internet de manière anonyme selon ces mêmes principes de virtualisation se développe.
  4. Les ressources témoignant d'un domaine culturel sont d'une très grande variabilité et hétérogénéité, leur identification et indexation comme support de contextualisation nécessitent une approche multimodale propice à la détection et suggestion de similarités. L'utilisation d'ontologies et de ressources sémantiques n'est pas possible car cela fixerait trop le contexte et il est probable que leur organisation serait constamment remise en cause. Les approches incrémentales où le système sollicite régulièrement le retour de l'utilisateur et s'adapte en conséquence sont à développer.
  5. La génération de texte explicatif à partir de données hétérogènes est un défi en soi. S'il n'existe pas de meilleur résumé possible, nous avons l'intuition de ce qu'est un bon résumé et sa lisibilité ainsi que son style sont essentiels. S'il est possible d'évaluer automatiquement l'informativité relative de plusieurs résumés, la lisibilité est bien plus suggestive.

---

## 5) Dimension interdisciplinaire

Notre projet s'inscrit dans l'axe « le numérique au service des arts et du patrimoine » de l'ANR et contribue prioritairement à l'objet plateforme éditoriale / Open Access avec:

- La mise en œuvre de protocoles et de méthodologies de traitements des données massives du web ;
- La transformation des données en connaissances ;
- La simulation numérique ;

Le projet concerne aussi l'axe « dimension sociale du tournant numérique » en ce qu'il aborde l'analyse des nouveaux modes de production des connaissances et de diffusion des informations.

---

## 6) Positionnement dans Agorantic

Ce projet est susceptible de trouver un prolongement espéré par les acteurs de l'économie numérique et des données partenaires d'Agorantic. Il a en effet comme perspective de contribuer aux différentes modélisations portant sur l'activité informationnelle dans les différents espaces ainsi que l'instrumentation proposée. Ces modèles ont vocation à fonder des outils de simulation numérique, permettant aux acteurs présents sur les réseaux d'optimiser leurs stratégies de positionnement et de diffusion d'information. Les enjeux sont à la fois sociétaux et économiques :

- *Axe 1 - Culture et numérique* : la mise à disposition d'outils favorisant l'extraction d'information à partir d'une masse de données hétérogènes, fluctuantes et non organisées contribue à la construction de connaissances et à leur appropriation sociale. Les enjeux se déclinent en termes de développement des usages (produsage) marchands, non marchands, d'éducation artistique et culturelle et d'information citoyenne.
- *Axe 2 - Réseaux sociaux, structures, contenus et usages* : les fonctions permettant d'inférer les structures sous-jacentes des productions informationnelles et des interactions d'acteurs sont un préalable à la proposition de valeur associée à la valorisation des données hétérogènes collectées à partir du web.

À titre d'exemple, la réussite à moyen terme de notre projet permettrait de proposer aux institutions culturelles des éléments méthodologique et logiciel pour mesurer, de manière indépendante l'impact potentiel des événements de type exposition, médiation, débat, hors les murs ou atelier participatif. L'hypothèse est ici de faire apparaître des entités collectives caractérisées dont la pérennité est plus longue que celle des individus aux identités plurielles et aux inscriptions circonstanciées.

---

## 7) Partenariats extérieurs

### *L'École de sciences de l'information (School of Information Studies - SIS) de l'Université McGill*

#### **Responsable pour le projet: Elaine Ménard**

L'école a pour mission de former des professionnels et des chercheurs à la gestion de l'information consignée et des connaissances en vue de maximiser son transfert et son utilisation. Nouvellement affiliée à la Faculté des Arts (Septembre 2014), SIS offre un programme de maîtrise conforme aux exigences de l'agrément de l'American Library Association (ALA), ainsi qu'un programme de doctorat en sciences de l'information. SIS met l'accent sur les connaissances et les compétences nécessaires à l'identification, l'acquisition, l'organisation, la recherche et la diffusion de l'information pour répondre aux besoins de la population en matière d'information diversifiée, des connaissances et des environnements d'apprentissage. SIS a également pour mission de contribuer à l'avancement des connaissances et des pratiques en gestion de l'information par ses activités de recherche aux plans national et international. Grâce à l'expertise variée des professeurs/chercheurs couvrant les multiples facettes des sciences de l'information, SIS se positionne comme l'un des leaders dans la société de l'information au Québec, au Canada et à l'étranger.

### *Institut de Recherche en Science de l'Information et de la Communication (IRSIC- EA 4262)*

#### **Responsables pour le projet: Fidelia Ibekwe**

Équipe de recherche en Sciences de l'Information et de la Communication associée à l'École de Journalisme et de Communication de l'Université d'Aix-Marseille (EJCAM), l'IRSIC est spécialisée sur trois pôles : médias (nouveaux médias, nouvelles pratiques journalistiques), Communication des organisations et Information (Extraction, Recherche d'information, enjeux de Big Data et de l'Open Data), Communication d'action et d'utilité sociétales.

Son expertise dans le projet se situe à la fois en amont et en aval : lors de l'identification des sources et supports d'objets informationnels à contextualiser et en aval pour assister les professionnels de musée dans la scénarisation, la mise en forme éditoriale des contenus à communiquer, afin que les modalités de la communication soient adaptées aux usagers cibles. L'IRSIC a une expertise reconnue en matière de communication engageante. Ainsi, elle mettra en œuvre des stratégies permettant d'amener les usagers des musées à participer à définir ce que devrait être leur musée et ce qu'il devrait transmettre, en établissant une relation directe avec les conservateurs qui évite l'intermédiaire "marketing" et les techniques publicitaires. Le conservateur avec l'aide du dispositif informatique communique sur le contexte culturel implicite au musée. L'utilisateur peut y répondre en apportant son propre témoignage ou suggestion. Les témoignages des usagers viennent ainsi alimenter les ressources de contextualisation.

---

## 8) Budget prévisionnel et financements envisagés

Demande Agorantic : 6000 €

**Cette somme doit permettre d'initier le projet et de mettre en œuvre la collaboration franco canadienne afin de répondre à l'appel ANR : Défi Société de l'information et de la communication/Coopération internationale Canada – NSERC pour avril 2015.**

Utilisation du budget :

- Indemnité et accueil stage master 2 informatique / réseaux : 2000 €
  - Invitation Elaine Ménard de l'Université McGill : 1000 €
  - Communication dans deux conférences internationales : 2000 €
  - Matériel et location de services informatique : 1000 €
-

## 9) Références bibliographiques

1. Julien Gaillard, Marc El-Bèze, Eitan Altman and Emmanuel Ethis : *Well-argued recommendation: adaptive models based on words in recommender systems*, EMNLP, October 18-21, Seattle, USA, 2013.
2. Romain Deveaud, **Eric SanJuan**, Patrice Bellot : *Estimating topical context by diverging from external resources*, SIGIR 2013: 1001-1004.
3. **Pierre Jourlin**, Romain Deveaud, **Eric SanJuan**, Jean-Marc Francony, Françoise Papa : *Design, implementation and experiment of a YeSQL Web Crawler*, SIGIR 2012 Workshop on Open Source Information Retrieval (OSIR 2012), Portland (USA), 16 August 2012.
4. Mathias (Mayeul) , Moussa (Assema), Zhou (Fen), **Torres-Moreno (Juan-Manuel)**, **Poli (Marie-Sylvie)**, Josselin (Didier), El-Beze (Marc), Carneiro Linhares (Andrea), Rigat (Françoise), *Optimisation using Natural Language Processing: Personalized Tour Recommendation for Museums*. Workshop C014, Computational Optimization 2014, Pologne.
5. **Marie-Sylvie POLI**. 2013. « Éducation et musée », p. 165-189, in *Culture & Musées*, Numéro Hors série pour les 20 ans de la revue, « La muséologie : 20 ans de recherches », Sous la dir. de Hana Gottesdiener et Jean Davallon, Actes Sud.
6. **Marie-Sylvie POLI**, Pascale ANCEL. 2014. *Exposer l'histoire contemporaine Évaluation muséologique d'une exposition : « Spoliés ! L' « aryanisation » économique en France 1940-1944*. Paris : Éd. La documentation française. (Musées-Mondes).
7. **Fidelia Ibekwe-SanJuan, Elaine Ménard** (sous dir.), *Les archives, bibliothèques et musées à l'ère du web social et participatif*, Numéro spécial *Revue Canadienne des Sciences de l'Information et des Bibliothèques*, à paraître 2015.
8. **Ibekwe-SanJuan F.**, *Big Data, Big machines, Big Science : vers une société sans sujet et sans causalité ?*, 19ème congrès de la SFSIC, Toulon 4-6 juin 2014, 9 pages
9. **Ibekwe-SanJuan F.**, Thomas Dousa (eds.), *Theories of Information, Communication and Knowledge. A Multidisciplinary approach*. Springer, Netherlands. Springer's Series Studies in History and Philosophy of Science, vol. 34, 2013, 331 pages
10. **Bernard Françoise** (2012), *Savoirs et actions : vers une culture de l'écoresponsabilité en Méditerranée*. In Bernard F & Durampart M (eds). *Savoirs en action, cultures et réseaux méditerranéens*, Paris : Hermès.
11. **Bernard Françoise** (2014), *Imaginaire, participation, engagement et empowerment*. Des notions pour penser la relation entre risques et changements. In Élisabeth Gardère et Nicole D'Almeida (coordination), *Risques mineurs, changements majeurs, Communication & Organisation*, n°45, 87-98.
12. C. Chen, **F. Ibekwe-SanJuan**, J. Hou., *The Structure and Dynamics of Co-Citation Clusters: A Multiple-Perspective Co-Citation Analysis*, *Journal of the American Society for Information Science & Technology*, 61(7), 1386–1409, 2010
13. **Juan-Manuel Torres-Moreno**: *Three Statistical Summarizers at CLEF-INEX 2013 Tweet Contextualization Track*. CLEF (Working Notes) 2014:565-573
14. **Juan-Manuel Torres-Moreno**, *Automatic Text Summarization*, Wiley & Sons, 2014.