



MINISTÈRE DE LA CULTURE ET DE LA COMMUNICATION

N°2014-01

FEUILLE DE ROUTE STRATEGIQUE

Métadonnées culturelles et transition Web 3.0

Ministère de la Culture et de la Communication
Secrétariat général
182, rue Saint-Honoré, 75033 Paris cedex 01

Feuille de route stratégique
Métadonnées culturelles et transition Web 3.0

JANVIER 2014

SOMMAIRE

Présentation de la feuille de route	5
Introduction	6
Action 1 Mettre en place un dispositif d'identifiants pérennes pour les ressources culturelles	11
Action 2 Développer un dispositif d'identification des auteurs de ressources culturelles	14
Action 3 Créer et maintenir une interconnexion sémantique des grands référentiels culturels : le "Graphe Culture"	17
Action 4 Expérimenter l'interconnexion inter-institutionnelle des données culturelles	20
Action 5 Explorer des modes d'interaction avec les publics utilisant les potentialités du Web 3.0	22
Action 6 Mettre en place un dispositif global de R&D destiné au soutien de l'innovation du Web 3.0 culturel	25
Action 7 Appliquer les technologies 3.0 à la description des événements culturels	27
Action 8 Positionner le Ministère de la Culture et de la Communication en tant qu'expert sur la traçabilité des données numériques	29
Action 9 Mettre en place un programme de sensibilisation et de formation interne pour le Ministère de la Culture et de la Communication	32
Annexes	34

PRESENTATION DE LA FEUILLE DE ROUTE

Cette feuille de route sur les métadonnées culturelles et la transition Web 3.0 du ministère de la Culture et de la Communication a été réalisée en prenant appui sur les actions 4 et 5 de la feuille de route stratégique ministérielle en faveur de l'ouverture et du partage des données publiques (« open data ») et s'inscrit dans le cadre des travaux sur le Web sémantique initiés par le Conseil d'Orientation de l'Édition Publique et de l'Information Administrative (COEPIA).

Cette feuille de route pose les premiers jalons d'une politique culturelle de transition vers le Web 3.0 intégrant pleinement la valeur des métadonnées culturelles. Elle n'a pas vocation à couvrir l'ensemble des aspects du 3.0 dans cette phase d'initialisation. Il s'agit de mettre en exergue des actions décisives et opérationnelles à mener par le secteur culturel à court et moyen terme dans l'agenda numérique 2014-2015. Afin de conduire les 9 actions identifiées dans la présente feuille de route, un groupe d'experts opérationnels dédié au pilotage de chaque action sera identifié. Ce groupe de travail spécialisé aura pour mission de définir techniquement le périmètre précis des livrables de chaque action et de mobiliser les experts du secteur pour s'assurer de la qualité scientifique des travaux conduits. Les membres du groupe de travail auront pour mission dans le calendrier déterminé en amont de piloter et de suivre la conception et la réalisation des projets définis.

Il est précisé que les expressions « Web 3.0 », « Web sémantique » et « Web de données » seront considérées par commodité de langage comme quasi équivalente dans les développements qui suivent.

Le département des programmes numériques qui a coordonné l'élaboration de cette feuille de route remercie chaleureusement les experts en Web sémantique de la Bibliothèque nationale de France (*Romain Weinz, Agnès Simon*), la Cité de la musique (*Rodolphe Bailly*), le Centre d'art et de culture Georges Pompidou (*Emmanuelle Bermès*), la direction générale des patrimoines (*Katell Briatte, Jean Davoigneau*), le service interministériel des archives de France (*Claire Sibille de Grimouard*), la direction générale de la création artistique (*Thierry Giacomino*), la direction générale des médias et des industries culturelles (*Florent Palluault, Anna Svenbro*), la délégation générale de la langue française et aux langues de France (*Thibault Grouas*), le Secrétariat général (*Marie-Véronique Leroi, Jean-Severin Lair*). Le département exprime en outre toute sa gratitude à Inria (*Fabien Gandon, Alexandre Monnin*) et aux experts du Web sollicités dans le cadre de ces travaux. Ces remerciements s'adressent également à Bertrand Sajus, Chef de projet et expert Web sémantique au département des programmes numériques qui a coordonné les travaux de ce groupe de travail.

INTRODUCTION

Depuis ses origines, le Web imaginé par Tim Berners Lee, dont les premiers balbutiements remontent à 1989, s'est transformé et a engendré des paradigmes successifs. Le Web 2.0 ou web collaboratif a marqué une étape majeure par son caractère participatif. Le Web sémantique a renforcé, d'une part, la « trouvabilité » sur Internet d'une information contextualisée et, d'autre part, la qualification des données et des relations entre elles pour classer et exploiter plus intelligemment l'information. L'Internet des objets qui est en train d'émerger permet quant à lui l'apparition d'une multitude d'objets de la vie quotidienne reliés à Internet. Il est donc aujourd'hui possible de parler de la coexistence de plusieurs Web. Dans le cadre des développements qui suivent nous nous intéresserons plus précisément au Web 2.0 et au Web sémantique dit « Web 3.0 ».

Le Web 2.0 associe divers systèmes sociaux complexes tels les blogs, wikis ou réseaux sociaux. Il fait référence à une certaine vision de l'Internet comme espace à part entière de socialisation. L'interaction entre pairs et la mise en relation avec autrui motivent l'action en tant que telle et assurent une production continue de contenus. Le public est dans ce contexte considéré comme producteurs, agrégateurs, partageurs ou lieux. Ce constat met en avant le rapprochement inévitable de la sphère de l'information et de la sociabilité.

Le Web sémantique qui fait de la sémantique des données - c'est à dire du sens attribué aux données - le pivot du Web, apparaît de plus en plus comme l'élément de fond du Web, utilisé et mis en valeur dans des sphères de plus en plus actives et visibles. Proposer une solution pour relier un ensemble d'informations non structurées à des nuages de données ouvertes et liées (« *linked open data* ») est devenu en très peu de temps un des grands chantiers du Web. L'enjeu, pour le Web sémantique, n'est pas ainsi de construire des plateformes de consultation de données, mais bien de rendre accessibles de l'information de référence dans des formats permettant leur consultation et leur diffusion dans les autres dimensions du Web comme le Web 2.0, tout en favorisant leur croisement et leur transformation. Les technologies du Web sémantique permettent dans ces conditions d'établir des liens entre des corpus autrefois isolés les uns des autres, voire d'autoriser de la sorte des découvertes de l'ordre de l'inattendu.

Au fur et à mesure que les techniques du Web se sont développées, le tropisme vers les données elles-mêmes n'a eu de cesse de s'accroître ; si bien que l'économie numérique qui se construit s'est structurée autour de l'échange de données, de masse de données ou *big data*.

Cet échange de données est bénéfique aussi bien pour la personne qui met à disposition que pour le réutilisateur. Le nouveau service, consommateur de données, gagne en utilité pour ses utilisateurs, et le système producteur voit sa position compétitive renforcée par la dépendance accrue du marché à ses services. Il ne faut pas oublier que la donnée isolée ne présente que peu d'intérêt en tant que telle. Sa véritable valeur dérive de son agrégation, de son recoupement, de son analyse et de la réutilisation qui en est faite.

Il est donc prévisible que la montée en puissance du Web 2.0 et du Web sémantique renforcent des possibilités d'échanges de données jusqu'alors insoupçonnées en permettant une classification et une structuration des contenus informationnels par la mobilisation de l'intelligence collective c'est-à-dire de la puissance créatrice et inventive des individus.

Dans ce contexte, il est primordial que le secteur culturel qui a une longue tradition en matière de structuration et de maintenance de référentiels reposant sur des métadonnées numériques, s'impose de manière plus offensive sur Internet. Les services et établissements sous tutelle du ministère de la Culture et de la Communication produisent des données publiques de qualité qui, si elles sont intelligemment interconnectées constitueront à terme un écosystème numérique de grande valeur culturelle qui deviendra le socle d'une économie culturelle

créative, inventive et partagée irriguant les territoires.

La souplesse et la puissance des technologies du Web sémantique dite technologies 3.0 transforment dans ces conditions le Web en une immense base de connaissance ouverte en perpétuelle évolution, et offrent ainsi des perspectives de développement stratégiques pour les institutions culturelles. Il suffit de penser aux potentialités offertes pour une institution patrimoniale de réussir à exprimer d'une manière formelle que tel monument a pour architecte telle personne dont la biographie est accessible dans telle base de données, que ce monument a été construit à telle date et est géolocalisé à tel endroit et visible sur telle carte. Le rebond de données en données organiquement liées entre elles constitue un véritable écosystème culturel vivant.

Pour autant, si certains établissements culturels français se sont positionnés parmi les pionniers mondiaux du Web sémantique, la prise de conscience des enjeux du « 3.0 » est encore très inégale dans la sphère publique.

Afin d'y remédier, la ministre de la Culture et de la Communication Aurélie Filippetti a fait de la problématique de la transition 3.0 du secteur culturel, l'un des leviers prioritaires de sa politique ministérielle des usages numériques initiée à l'occasion de l'Automne numérique.

Investir sur des technologies de rupture est ainsi devenu une action déterminante pour continuer à faire du secteur culturel un secteur d'avenir qui sait imposer ses valeurs sur Internet. En ce sens, le travail sur l'enjeu des métadonnées culturelles ou ensemble de données structurées décrivant des ressources culturelles physiques ou numériques, est primordial pour l'avenir des politiques culturelles et participe au rayonnement de la culture française et de la francophonie sur Internet.

Dans cette perspective, un certain nombre de questions se posent : dans quelle mesure les institutions culturelles doivent-elles et peuvent-elles développer des stratégies de développement trouvant leur fondement dans l'*open data*, le crowdsourcing, l'indexation contributive ou plus généralement la collaboration avec leurs publics ? Comment favoriser l'émergence de nouvelles pratiques artistiques et culturelles ainsi que le développement de nouveaux modèles économiques dynamiques prenant en compte la puissance inventive de l'intelligence collective ? Dans quelle mesure l'identification des objets, leurs modélisations, leurs interconnexions sémantiques augmentent l'interaction entre le Web et le monde tangible de la vie quotidienne ? Comment la richesse d'indexation du patrimoine culturel historique peut-elle profiter à tous les citoyens en étant complétée par des métadonnées exploitées par un écosystème numérique de la création et de l'innovation soutenu par la puissance publique ?

Ces interrogations qui sont au cœur des nouveaux systèmes d'organisation et de représentation des données sur le Web ainsi que des modes d'accès à l'information en voie de diversification, orchestrés par les métadonnées, sont loin d'être anecdotiques. Elles posent de manière sous-jacente la question de la confiance, clé de voûte du Web, qu'il convient d'accorder ou pas à ces systèmes basés sur les données.

A l'heure où le Web imprègne de plus en plus les actes de notre quotidien, investit les stratégies de développement des acteurs publics et privés, participe au renouvellement et à la modernisation des politiques publiques et irrigue chacun des secteurs de nos économies modernes, il est important que le secteur public et pour ce qui nous concerne le secteur culturel joue un rôle décisif dans l'établissement de la confiance que les citoyens accorderont à un Web de la Culture.

Prendre en compte ce concept de « confiance » dans le pilotage et l'élaboration des politiques publiques est donc crucial.

L'engagement institutionnel n'est plus seulement « éditorial » au sens d'un effort que les institutions se doivent de fournir pour garantir la véracité et la qualité du contenu des documents qu'elles mettent en ligne ; mais concerne tout autant les processus de co-production, de co-design, de structuration, d'ouverture et de partage de données indissociables du substrat technologique que constitue le Web. Dans ce nouveau paradigme qui est en train de se construire, les institutions publiques doivent donc s'investir en amont et en aval d'une simple mise en ligne de documents.

Les institutions publiques culturelles ont à jouer une place déterminante dans la production et la coproduction des biens communs de la connaissance. Elles sont concernées aussi bien par le dispositif mondial d'identification des auteurs d'œuvres culturelles, que par l'interconnexion fiable des bases de connaissances détenues ou produites par les services, établissements ou organismes culturels, ou encore par le développement d'expertise en matière de traçabilité des métadonnées.

La présente feuille de route identifie 9 actions opérationnelles permettant au secteur culturel de prendre en compte les opportunités et les défis qu'offrent le 3.0. Les développements de cette feuille de route s'inscrivent dans la lignée de la politique volontariste d'ouverture et de partage des données publiques culturelles, et en complément des études de faisabilité en cours dans le cadre des travaux de la direction générale des médias et des industries culturelles sur la possibilité de créer un portail d'identification des œuvres et des ayants droit.

Camille Domange
Chef du département des programmes numériques

ACTIONS

ACTION 1 METTRE EN PLACE UN DISPOSITIF D'IDENTIFIANTS PERENNES POUR LES RESSOURCES CULTURELLES

RESUME

L'identification normée et pérenne des ressources accessibles par Internet est un élément essentiel du Web sémantique. Il convient de mettre progressivement en place un dispositif d'identification pérenne, adapté aux besoins des institutions culturelles.

DESCRIPTION

L'interopérabilité du Web et son existence en tant qu'espace d'information global et unifié reposent sur l'existence de normes communes, en particulier sur une norme d'identification. Les URL (*Uniform Resource Locator* – localisateur uniforme de ressource) sont les adresses que nous utilisons tous les jours pour accéder à toutes les ressources du Web. Ces URL sont bien davantage que des adresses : ce sont des identifiants universels qui fondent l'existence des ressources sur le Web.

De la même manière, dans le Web sémantique, toute ressource doit bénéficier d'un identifiant qui respecte une syntaxe particulière et qui peut être actionné pour accéder à la ressource. Ces identifiants sont les URI (*Uniform Resource Identifier* - identifiant uniforme de ressource). Ils partagent la même syntaxe que les URL mais avec de surcroît une exigence forte de pérennité. Sans URI, aucune ressource ne peut être produite ou utilisée dans le Web des données liées (*Linked Data*) : cela concerne aussi bien les ressources numériques (une image, un livre numérisé, une vidéo, etc.) que les entités réelles (objets, personnes, lieux, institutions, etc.) ou abstraites (concepts, sujets, thèmes, classifications) dès lors qu'on élabore la description de ces ressources.

L'attribution de ces identifiants ou URI pose principalement deux problèmes : il faut être capable de garantir leur unicité au moment où on les crée et il faut pouvoir garantir leur pérennité sur le long terme.

L'unicité se définit comme une unicité globale, qui se vérifie sur l'ensemble des ressources quelle que soit leur origine géographique ou leur nature. La pérennité est quant à elle plus complexe à définir, car elle nécessite de déterminer si l'identifiant doit avoir une longévité équivalente à la ressource ou supérieure. Par exemple, un objet patrimonial dont la charge passe d'une institution à une autre doit-il recevoir un nouvel identifiant ? Qu'advient-il de son ancien identifiant ? Etc.

Diverses solutions techniques peuvent être mises en place pour répondre à ces deux exigences d'unicité et de pérennité. Les organisations qui souhaitent produire et partager des références de ressources durables, notamment dans une perspective patrimoniale, ont besoin d'un système d'identification dont la maintenance et la continuité de service sont garanties à très long terme.

Toutefois, il s'agit moins aujourd'hui d'un problème technique qu'organisationnel. En effet, les outils actuels fournissent les moyens de pérenniser les URL et les URI, mais cela requiert un effort de gestion et de maintenance qui peut représenter un coût (par exemple, lorsqu'on change l'outil de production d'un site Web institutionnel, il faut prévoir la mise en place du système de continuité des identifiants). Les institutions qui produisent des URI doivent donc disposer des moyens nécessaires pour mener à bien cette mission. De plus, elles doivent

s'inscrire dans un dispositif global de suppléance au cas où elles seraient dans l'incapacité de continuer à maintenir les identifiants qu'elles ont créés.

La création et la maintenance des URI pérennes relèvent donc autant de l'organisation interinstitutionnelle que de la technologie. Il est nécessaire que soient partagées au niveau national les bonnes pratiques qui président à la garantie de l'unicité et de la pérennité des URI, ainsi que des recommandations de solutions techniques diversifiées et adaptées aux différents contextes institutionnels : le ministère de la Culture et de la Communication ainsi que ses établissements publics sous tutelle.

Les recommandations devront prendre en compte la nature des différentes ressources à identifier et l'existence d'URI déjà attribuées, dans certains cas, par des institutions ou des organismes au niveau national ou international. Ces bonnes pratiques et recommandations constitueront un cadre qui devra être complété par un plan de suppléance pouvant être appliqué au cas où une institution chargée de maintenir des URI viendrait à disparaître ou à être modifiée significativement d'une façon qui affecte les objets dont elle a la charge ou leurs identifiants (changement de nom, évolution des missions, division d'une institution en plusieurs entités, etc.). Afin de garantir la cohérence, la fiabilité et l'indépendance de ce dispositif ainsi que son adaptation aux besoins des institutions culturelles, il convient de l'élaborer au plan national en tenant compte du contexte international et des travaux du W3C.

ENJEUX STRATEGIQUES

Fournir aux institutions culturelles un cadre de bonnes pratiques et de recommandations techniques pour les accompagner dans la création et la maintenance de leurs URI.

Assurer au plan national l'inscription dans les missions et objectifs des institutions d'une fonction consistant à garantir l'unicité et la pérennité des URI.

Proposer aux institutions qui n'ont pas les moyens d'offrir ces garanties un plan de suppléance pour l'attribution et la maintenance de leurs URI.

Assurer une période de transition pendant laquelle la situation ne sera pas homogène, entre les différentes institutions, sur le plan de la mise en place des URI.

RISQUES A NE PAS FAIRE

Les institutions qui ne peuvent pas produire des identifiants uniques et pérennes seront dans l'incapacité d'utiliser les technologies du Web sémantique, ces dernières reposant sur les URI.

Si des identifiants ne sont plus maintenus et deviennent obsolètes, l'ensemble des services construits sur les données concernées cesseront de fonctionner.

POINTS D'ATTENTION

Le dispositif mis en place doit favoriser une logique de convergence tout en préservant une certaine souplesse dans les méthodes d'implémentation, qui doivent être adaptées aux besoins des différentes institutions.

Il est important d'adopter une définition pragmatique de la notion de pérennité afin de ne pas entraîner de surcoûts dans la mise en œuvre. La pérennité d'un service doit être proportionnée

aux besoins et aux moyens.

Il est nécessaire de fournir un cadre informatif aux institutions qui pourront ainsi évaluer les standards et systèmes d'identification existants et adaptés à leurs ressources numériques.

Les institutions culturelles créent et maintiennent déjà de nombreux systèmes d'identifiants métiers, sans rapport avec le Web (cotes des bibliothèques, n° d'inventaire des musées, etc.) Ces identifiants métiers doivent être pris en compte dans la logique de création des URI.

LIEN AVEC D'AUTRES ACTIONS

Cette action s'inscrit en lien avec l'ensemble des autres actions de la feuille de route.

LIVRABLES DU GROUPE DE TRAVAIL SPECIALISE

État de l'art sur la question des identifiants pérennes au plan international, en collaboration avec Inria dans le cadre d'un programme de R&D (*cf. action n° 6*).

Vade-mecum technique, à l'attention des producteurs de données, pour l'adoption progressive d'identifiants pérennes.

Note de recommandation pour l'installation d'un comité permanent d'experts sur la question des identifiants pérennes.

ACTION 2 DEVELOPPER UN DISPOSITIF D'IDENTIFICATION DES AUTEURS DE RESSOURCES CULTURELLES

RESUME

L'identification des auteurs est indispensable à la diffusion des ressources culturelles sur le Web. L'identifiant international ISNI (*International Standard Name Identifier*) est une opportunité pour le partage des données sur les auteurs dans le secteur culturel et pour le rayonnement des auteurs français.

DESCRIPTION

Les métadonnées d'identification constituent le socle du Web des données liées. Celles qui concernent les auteurs d'œuvres de l'esprit sont cruciales, car elles sont indispensables à la traçabilité des œuvres, leur visibilité, leur compréhension, leur conservation, leur partage, ainsi qu'à toute l'économie du numérique.

Le mot « auteur » est entendu au sens de personnes physiques et morales prenant part à la constitution d'une ressource, qu'elles soient créateurs, contributeurs ou interprètes, dans tous les domaines culturels : patrimoine, industries culturelles et création artistique contemporaine.

Par leurs missions, les institutions gèrent déjà traditionnellement des référentiels sur des auteurs, locaux ou nationaux, spécialisés ou encyclopédiques. Certaines souhaitent développer la création et la consolidation de référentiels spécialisés, lorsqu'ils sont inexistantes. Or, les institutions culturelles ont besoin d'identifier de manière fiable des identités : désambiguïsation des homonymes, distinction des identités publiques d'une même personne dans les cas de pseudonymie, relations entre auteurs. Par ailleurs, elles ont besoin d'identifier des créateurs ayant travaillé dans des secteurs différents et de partager les données produites dans ces différents contextes.

Aujourd'hui, l'essor du Web et la mondialisation augmentent les exigences d'identification et les situent sur un plan international. C'est dans ce cadre qu'a été créé, en 2011, un dispositif mondial d'identification : l'*International Standard Name Identifier* (ISNI). La base ISNI compte aujourd'hui près de 7 millions d'identifiants sur des auteurs contemporains et patrimoniaux.

L'ISNI, conçu pour identifier de manière unique et pérenne tout contributeur de contenus culturels, scientifiques et techniques à travers le monde, représente un avantage sans précédent. L'identifiant repose sur une norme internationale (ISO 27729:2012). Il est gouverné par l'agence ISNI-IA (*ISNI-International Agency*), société de droit britannique basée à Londres, qui est l'autorité d'enregistrement ISO de l'identifiant. L'ISNI-IA réunit des acteurs publics (bibliothèques) mais aussi privés (sociétés de gestion de droits, éditeurs). La Bibliothèque nationale de France, en tant que co-fondatrice de ce dispositif, est membre du comité des directeurs de l'ISNI-IA ; elle y représente la CENL – Conférence des directeurs des bibliothèques européennes.

L'identification internationale, unique, pérenne et fiable que propose l'ISNI est indispensable à l'échange et à la diffusion des données entre différents secteurs :

- Elle favorise le rapprochement et l'interopérabilité des données des institutions culturelles. L'ISNI servira de clé unique de connexion entre les données des administrations, des musées, bibliothèques, archives, émanant de ou portant sur les

personnes morales créatrices ou détentrices d'archives papiers ou numériques. En outre, les institutions culturelles françaises possèdent des données spécifiques et sectorielles susceptibles de servir de sources de référence pour la fiabilisation des données de la base centrale ISNI, en fonction de leurs domaines d'expertise. La coopération entre ces organismes est donc un facteur majeur de la réduction des coûts de production des données.

- Elle permet le rapprochement entre les données culturelles des secteurs publics et privés. Le rapport Lescure (proposition 79) a souligné la nécessité de relier les métadonnées issues de tous les acteurs de l'interprofession dans l'objectif d'une meilleure gestion des droits. Le rapport cite ainsi l'ISNI parmi les métadonnées de propriété visant « à faciliter la circulation des métadonnées et à éviter les problèmes d'identification liés à l'orthographe ou à l'homonymie ». A l'échelle européenne, la commission européenne, recommande l'emploi d'identifiants uniques pour la gestion des droits, en particulier dans le secteur musical (*Brussels, 24.5.2011. COM(2011) 287 final*).

- Elle constitue un pivot entre les données culturelles et permet leur diffusion sur le Web de données (*cf. action n° 3*). En effet, l'ISNI est publié sous la forme d'un lien fiable et pérenne sur le Web ou URI (*Uniform Resource Identifier, cf. action n°1*), en conformité avec les standards du Web. L'ISNI favorise ainsi la création de nouveaux services de découverte et de navigation intuitive entre les ressources culturelles sur Internet.

Enfin, promouvoir cet identifiant répond à des enjeux de rayonnement des auteurs de la création contemporaine émergeant en France et, par-là même, du patrimoine culturel français sur le Web de demain. Afin de couvrir les différents champs culturels en France et d'assurer au meilleur coût le meilleur équilibre entre la complétude, la rapidité et la fiabilité du processus d'identification au plan national, il est nécessaire d'une part, de renforcer la collaboration entre les institutions culturelles autour de la production de référentiels et de l'échange de données convergeant vers l'ISNI, cadre d'identification à long terme des auteurs ; et d'autre part, de développer une stratégie nationale de promotion, d'adoption et de diffusion d'ISNI comme identifiant international de référence.

ENJEUX STRATEGIQUES

Faciliter la gestion des droits des auteurs contemporains.

Accroître la visibilité des données patrimoniales et de la création contemporaine française à l'international.

Développer la collaboration des différents acteurs du domaine culturel pour le rapprochement de référentiels variés.

Engager le rapprochement des données produites ou détenues par le secteur public avec les données du secteur privé.

Favoriser la création de services de découverte des ressources culturelles sur le Web et de navigation intuitive.

Réduire les coûts de production des données.

RISQUES A NE PAS FAIRE

Exclusion des données culturelles la scène internationale.

Éclatement progressif des systèmes d'identification.

POINTS D'ATTENTION

Mettre en place un modèle économique garantissant l'attribution et la diffusion de ces identifiants pour les institutions publiques, détentrices de bases de données sur les auteurs.

Apprendre aux acteurs du secteur culturel à évoluer en intégrant les besoins propres à différents secteurs d'activité.

Capitaliser sur des initiatives de productions et de consolidation de référentiels spécialisés pour les auteurs n'ayant pas été référencés.

Se lier aux identifiants locaux ou spécifiques (SIRET, SIREN, ORCID).

LIEN AVEC D'AUTRES ACTIONS

Action 1 : Mettre en place un dispositif d'identifiants pérennes pour les ressources culturelles

Action 3 : Créer et maintenir une interconnexion sémantique des grands référentiels culturels : le « Graphe Culture »

LIVRABLES DU GROUPE DE TRAVAIL SPECIALISE

État des lieux : panorama des données existantes sur les auteurs de ressources culturelles produites par les institutions culturelles françaises et leurs partenaires des industries culturelles.

Préconisations et méthodologie pour l'adoption et la diffusion des identifiants ISNI à une échelle nationale.

Identifier des projets pilotes sont destinés à tester l'échange interinstitutionnel de données d'identification d'auteurs convergeant vers l'ISNI.

ACTION 3 CREER ET MAINTENIR UNE INTERCONNEXION SEMANTIQUE DES GRANDS REFERENTIELS CULTURELS : LE « GRAPHE CULTURE »

RESUME

Les institutions culturelles détiennent ou produisent, depuis longtemps, de grands corpus de métadonnées structurants dans leurs divers domaines de connaissance. L'interconnexion de ces référentiels au moyen des technologies du Web sémantique augmentera à terme la valeur d'usage des ressources culturelles.

DESCRIPTION

Les « référentiels » sont des corpus de métadonnées structurants pour un domaine de connaissance. Il s'agit par exemple d'index permettant d'identifier des auteurs, de nomenclatures de lieux géographiques, de thésaurus spécialisés, de dictionnaires multilingues, etc.

Certains grands référentiels produits ou soutenus de longue date par les institutions culturelles françaises sont appelés à jouer un rôle majeur dans l'écosystème mondial du Web des données liées, lequel deviendra le « Web de la connaissance » de demain.

A titre d'exemples :

- RAMEAU contient plus de 167 000 concepts, noms de lieux et de personnes utilisés pour l'indexation des ouvrages de la BnF et de plusieurs centaines de bibliothèques françaises ou francophones ;
- Wikipédia contient plus de 1 300 000 articles encyclopédiques (disponibles dans le Web des données liées via Dbpédia en français) ;
- ISNI contient les éléments d'identification consolidés d'un million d'auteurs d'œuvres culturelles ;
- MIMO contient un thésaurus de 5 000 noms d'instruments de musiques traduits en 6 langues et reliés à plus de 50 000 ressources numériques muséales ;
- GINCO contient de nombreux thésaurus thématiques utilisés par le ministère de la Culture et de la Communication, qui couvrent des domaines aussi variés que l'architecture et l'urbanisme, l'archéologie, l'iconographie, les objets mobiliers, etc. Ces vocabulaires décrivent des millions de ressources numériques disponibles dans les bases de données culturelles.

Ces enjeux ont déjà été intégrés dans les stratégies de certains grands acteurs commerciaux du Web qui développent, depuis quelques années, leurs propres graphes - c'est-à-dire l'interconnexion sémantique des jeux de métadonnées selon des formats spécifiques - à cette technologie de connaissance et les utilisent pour augmenter la pertinence de leurs services.

La réalisation du graphe des institutions culturelles et son interconnexion fine avec les autres graphes en cours de constitution augmentera la visibilité et l'indexation par les moteurs de recherche des ressources et des données du ministère de la Culture et de la Communication et de ses établissements publics, jusque là cachées et inexploitées.

Cette interconnexion permettra du même coup de désenclaver les bases de données auxquelles les référentiels sont nativement liés (soit des millions de ressources culturelles). Elle permettra, par exemple, de retrouver automatiquement et avec certitude l'exhaustivité des

œuvres d'un auteur décrites dans une multitude de bases de données, ou encore de trouver de la documentation sur un événement historique décrit sous diverses appellations dans des sites Internet de différentes langues (Par exemple, Grande Guerre, Première Guerre mondiale, Guerre 14-18, World War I, WWI, Erster Weltkrieg, etc.).

L'interconnexion favorisera une meilleure économie dans la production de données et la création de nouveaux services.

Dans ce contexte, les données seront de moins en moins dupliquées mais de plus en plus liées, chaque producteur pouvant alors concentrer ses efforts sur sa propre valeur ajoutée. Cela favorisera l'émergence d'opérateurs tels que, par exemple, des méta-moteurs spécialisés. En tant que graphe de connaissance, les référentiels interconnectés seront donc utilisés bien au-delà de l'usage pour lesquels ils avaient été initialement conçus, générant ainsi, par effet de seuil, une très forte valeur ajoutée. Mais cette valeur dépendra en grande partie de l'étendue et de la fiabilité du graphe et son usage ne se développera qu'à la mesure de la confiance que lui accorderont les internautes. Pour être fiable, l'interconnexion sémantique des référentiels ne peut être réalisée par des moyens exclusivement automatiques. Elle requiert ainsi l'implication des experts qui en sont les éditeurs intellectuels. Au-delà de la phase d'initialisation, il convient également de prévoir la maintenance de l'interconnexion des référentiels, car ces derniers évoluent chaque jour.

Afin de réaliser le graphe de connaissance des grands référentiels culturels, qu'on pourrait appeler le « Graphe Culture », il est nécessaire :

- de créer un réseau de collaboration dédié, entre producteurs des référentiels, qui se donnera une méthodologie et des bonnes pratiques spécifiques au liage inter-référentiels ;
- de développer des outils informatiques mutualisés pour la création et la maintenance des liens inter-référentiels (technologie dite de « réconciliation de données » etc.) de manière à optimiser le travail des experts et automatiser tout ce qui peut l'être ;
- de mettre en place un dispositif garantissant une bonne disponibilité technique de ce graphe dans le Web des données liées, pour les humains et les machines (points d'accès, rafraîchissement des données, etc.).

Ce chantier, par nature inter-institutionnel, requiert un soutien et un accompagnement au plan national.

ENJEUX STRATEGIQUES

Désenclaver d'immenses bases de données constituées de millions de ressources numériques dont la description repose sur les grands référentiels culturels.

Offrir aux usagers une base de connaissances ouverte et réutilisable, y compris pour des usages commerciaux.

Produire un écosystème vivant de métadonnées propice à l'innovation technologique (technologie de recherche d'information, de visualisation, de navigation etc.).

Mettre en valeur, sur le plan international, la production des institutions culturelles françaises (richesse, fiabilité, confiance) *via* le Web des données liées.

RISQUES A NE PAS FAIRE

A une période clé de l'évolution technologique du Web, manquer l'occasion de renouveler puissamment les usages du riche patrimoine des référentiels culturels développé depuis plus de 30 ans par le ministère de la Culture et de la Communication et ses établissements publics.

Affaiblir la position des institutions culturelles françaises dans l'écosystème mondial émergent du Web des données liées, ou « Web de la connaissance ».

POINTS D'ATTENTION

L'inter-institutionnalité peut devenir excessivement complexe si l'on n'adopte pas une approche modulaire et progressive dans le déploiement du « Graphe Culture ».

En tant que méta-référentiel, le « Graphe Culture », n'a de sens que s'il s'inscrit dans une politique orientée développement durable, cela implique une prise en compte des contraintes de maintenance à long terme.

LIEN AVEC D'AUTRES ACTIONS

Action 2 : Développer un dispositif d'identification des auteurs de ressources culturelles

Action 4 : Expérimenter l'interconnexion inter-institutionnelle des données culturelles

Action 5 : Explorer des modes d'interaction avec les publics utilisant les potentialités du Web 3.0

Action 6 : Mettre en place un programme de R&D destiné à soutenir l'innovation du Web 3.0 culturel

LIVRABLES DU GROUPE DE TRAVAIL SPECIALISE

Sélection d'un premier groupe de référentiels susceptibles de constituer le noyau initial du « Graphe Culture »

Identifier un projet de liage sémantique des référentiels constitutif du noyau initial du « Graphe Culture »

ACTION 4 EXPERIMENTER L'INTERCONNEXION INTERINSTITUTIONNELLE DES DONNEES CULTURELLES

RESUME

L'interconnexion inter-institutionnelle des données est un enjeu essentiel du Web des données liées. Afin de sensibiliser les publics et les institutions à cet enjeu, il convient de finaliser ou réaliser des outils de démonstration sous forme de « preuves de concepts ».

DESCRIPTION

Le Web des données liées et les technologies du Web sémantique restent aujourd'hui abstraits et très peu connus du grand public y compris pour certains professionnels de la culture.

Afin d'explicitier de manière simple et pragmatique les enjeux de cette évolution, notamment sur le plan inter-institutionnel, il est aujourd'hui nécessaire de réaliser des projets sous forme de « preuves de concepts ».

L'approche « preuve de concept » est une démarche orientée utilisateur final. Elle permet de mettre en œuvre, à titre exploratoire, de nouveaux types de fonctionnalités et d'obtenir un premier retour d'expérience sur le plan des outils, des usages et des pratiques culturelles.

Ces preuves de concept devront être conçues pour leur caractère intuitif et démonstratif. Elles devront s'inscrire dans une démarche de désenclavement inter-institutionnel des données culturelles. Le ministère de la Culture et de la Communication ainsi que plusieurs grandes institutions telles que la Bibliothèque nationale de France (BnF), la Cité de la musique ou le Centre d'art et de culture Georges Pompidou ont déjà entrepris le traitement de leurs bases de données en vue de les publier sur le Web de données. Mais ces réalisations sont encore très centrées sur les ressources propres à chaque institution. Aucune expérience explicitement orientée grand public n'a encore été réalisée sur la base d'une interconnexion de corpus produits par différents établissements publics ou par le ministère de la Culture et de la Communication.

Or, l'un des enjeux les plus forts du Web des données liées est non seulement l'interconnexion de bases de données internes à chaque institution, mais aussi et surtout l'interconnexion profonde entre des données produites par des institutions différentes. Une telle approche génère de nouvelles connaissances et de nouveaux usages du seul fait du « tissage » fin de données issues de différents domaines du savoir.

En 2013, un premier pas a été franchi dans ce sens par la Cité de la musique et la BnF, qui ont mené conjointement une expérience d'interconnexion automatisée d'une partie de leurs données sur les compositeurs et interprètes (environ 15 000 personnes) communs aux deux établissements.

La prochaine étape consiste à affiner le traitement sémantique des données et à développer une interface de navigation grand public. Le panel des usagers est très large : de l'internaute curieux qui écoute des extraits d'œuvres au musicologue qui cherche à recouper des informations en provenance de ces différentes bases.

La démarche initiée par la Cité de la musique et la BnF s'inscrit pleinement dans la perspective

de l'interconnexion inter-institutionnelle des données. Elle éclaire par là-même les enjeux fondamentaux du Web des données liées. Ce projet exemplaire doit donc être finalisé dans un proche avenir et faire l'objet d'une communication grand public. Compte tenu de l'importance de ces enjeux, d'autres projets du même type, impliquant différents services du MCC et EP, doivent être soutenus au plan national afin de renforcer cette démarche.

ENJEUX STRATEGIQUES

Créer une dynamique entre les établissements pour le déploiement du Web des données liées

Démontrer que l'interconnexion des données améliore l'accès et augmente la visibilité des ressources qu'elles décrivent

Rationaliser et mutualiser l'usage des métadonnées entre établissements

RISQUES A NE PAS FAIRE

Laisser s'installer une réticence des institutions à s'engager dans cette nouvelle voie faute d'une vision concrète des principes et des enjeux

POINTS D'ATTENTION

Produire des preuves de concept ayant un impact sur un large public (éviter l'effet de niche en ne s'adressant qu'à des spécialistes)

Exploiter les résultats de ces expérimentations dans la mise en œuvre de projets pérennes

LIEN AVEC D'AUTRES ACTIONS

Action 3 : Créer et maintenir une interconnexion sémantique des grands référentiels culturels : le « Graphe Culture »

Action 9 : Mettre en place un programme de sensibilisation et de formation interne à l'attention des services du ministère de la Culture et de la Communication et des établissements publics sous tutelle

LIVRABLES DU GROUPE DE TRAVAIL SPECIALISE

Sélection de preuves de concepts à réaliser (ou finaliser) en tant que projets pilotes

Rédaction d'une note de cadrage pour chacun des projets sélectionnés

ACTION 5 EXPLORER DES MODES D'INTERACTION AVEC LES PUBLICS UTILISANT LES POTENTIALITES DU WEB 3.0

RESUME

Le Web 3.0 doit être envisagé comme une augmentation des aspects collaboratifs du Web, en continuité avec le 2.0. Il convient d'analyser de manière prospective l'impact social du Web 3.0, notamment sur le positionnement des institutions culturelles vis-à-vis de leurs publics.

DESCRIPTION

Le Web des données liées et les technologies du Web sémantique, qu'on appelle communément le « Web 3.0 », ne remettent aucunement en cause les acquis du Web 2.0. Les référentiels sémantiques n'auront de sens et de viabilité que s'ils sont produits et réutilisés dans le cadre du Web social ou collaboratif.

Au cours des années 2000, de nombreuses institutions culturelles se sont appropriées des pratiques 2.0 et ont inventé des modes d'interaction avec leurs publics tirant parti de ce contexte d'usage, notamment grâce aux réseaux sociaux. Les technologies sémantiques apportent une nouvelle génération d'outils. Ces outils offrent des modes d'interaction à la fois puissants et fins pour produire, modéliser et partager des connaissances. Les technologies 3.0 permettent non seulement de traiter l'information à un niveau de granularité très fin, la donnée, mais aussi de représenter les relations logiques entre ces données de telle sorte qu'elles soient interprétables par des machines.

En exposant leurs données ainsi traitées (données « sémantisées ») via des plates-formes 3.0, les institutions culturelles pourront interagir d'une manière profonde avec toute forme de communautés d'utilisateurs (club d'amateurs, sociétés savantes, associations professionnelles etc.). Elles pourront ainsi coproduire des corpus vivants, beaucoup plus complexes que ce que permettent les outils 2.0 popularisés dans les années 2000 (blogs, wikis de première génération, etc.). Elles fourniront aux experts externes les moyens techniques d'exprimer et d'interconnecter leurs savoirs avec un niveau de précision et d'expressivité inégalé. Ce sont de véritables bases de connaissances, voire des systèmes experts, qui pourront être ainsi cocréés en mode collaboratif ouvert.

Il existe déjà des démarches collaboratives prometteuses préfigurant cette évolution, même si elles n'exploitent pas encore les potentialités offertes par l'*open data* et les technologies du Web 3.0. C'est le cas, par exemple, du projet mené par les Archives départementales de la Vendée pour la co-construction de dictionnaires d'histoire locale. Ce projet a permis la réalisation d'une plate-forme constituée par l'ensemble de l'offre numérique des Archives départementales de la Vendée ainsi que par des dictionnaires s'appuyant sur le dépouillement des inventaires et archives en ligne auxquels ils font systématiquement référence, et se citant entre eux autant que possible.

Dans les prochaines années, les institutions seront amenées à se positionner à la croisée des communautés expertes, des collections patrimoniales et des pratiques culturelles auxquelles elles sont vouées, dans la perspective du développement du « Web de la connaissance ».

Afin de préparer cette évolution de positionnement des institutions culturelles, il est nécessaire de développer une analyse prospective des aspects sociaux du Web 3.0 et de définir

des actions transverses destinées à favoriser les pratiques sociales dans ce contexte technologique. Ces actions peuvent, par exemple, concerner la coproduction interactive de bases de connaissance liées à des lieux, des événements, des collections patrimoniales, etc.

ENJEUX STRATEGIQUES

Fidéliser et impliquer les usagers en suscitant leur participation par l'ouverture et le partage de données publiques

Développer la création de référentiels en coproduction avec les usagers

Faciliter l'échange des données entre les institutions culturelles et les usagers

Valoriser les compétences et connaissances des usagers

Favoriser le rayonnement du Web de données culturel français au plan mondial

RISQUES A NE PAS FAIRE

Sous-estimation de la dimension collaborative et sociale du Web 3.0 au profit d'une approche purement techniciste

Retard du ministère de la Culture et de la Communication et de ses établissements publics sous tutelle dans leur repositionnement institutionnel

POINTS D'ATTENTION

Penser les effets sociologiques de la transition vers le Web 3.0 non seulement du point de vue de l'utilisateur individuel, mais aussi du point de vue des communautés et des institutions culturelles

LIEN AVEC D'AUTRES ACTIONS

Action 3 : Créer et maintenir une interconnexion sémantique des grands référentiels culturels : le "Graphe Culture"

Action 4 : Expérimenter l'interconnexion inter-institutionnelle des données culturelles

Action 7 : Appliquer les technologies sémantiques à la description des événements culturels

LIVRABLES DU GROUPE DE TRAVAIL SPECIALISE

Rédaction d'un cahier des charges pour le lancement d'une étude comparative et prospective sur les aspects sociaux de la transition vers le Web 3.0. Les résultats de cette étude devront être fournis avant la fin du premier trimestre 2015

Publication d'une note de cadrage destinée à la formulation de propositions d'actions transverses destinées à favoriser les pratiques sociales du Web dans le contexte technologique 3.0, à l'attention des services du ministère de la Culture et de la Communication et de ses établissements publics

ACTION 6 METTRE EN PLACE UN PROGRAMME DE R&D DESTINE AU SOUTIEN DE L'INNOVATION DU WEB 3.0 CULTUREL

RESUME

La recherche et développement (R&D) est indispensable à l'essor des technologies du Web sémantique dans le secteur culturel. Il est aujourd'hui nécessaire de renforcer et développer globalement le cadre collaboratif initié en 2012 par la convention Sémanticpédia.

DESCRIPTION

Les projets exploratoires évoqués dans cette feuille de route impliquent une démarche résolument innovante. Au-delà des actions ponctuelles, il est indispensable de soutenir, à long terme, la dynamique de l'innovation par la mise en place d'une politique de recherche et développement (R&D) adaptée à l'essor du Web sémantique dans le domaine culturel.

Le ministère de la Culture et de la Communication dispose déjà d'un premier cadre de R&D, depuis novembre 2012, avec la convention Sémanticpédia [Cf. Annexe] instaurant un partenariat stratégique entre le ministère, l'association Wikimedia France et l'établissement public de recherche Inria. Au sein d'Inria, cette convention mobilise en particulier l'équipe Wimmics, mondialement reconnue pour son expertise dans le domaine des technologies du Web sémantique.

Ce partenariat a permis notamment la réalisation d'une extraction des contenus de Wikipédia à partir de sa version francophone se traduisant par une base de données conforme aux normes du Web sémantique : DBpédia en français. Cette base de données contient des informations non seulement en français mais aussi dans la plupart des autres langues majoritairement présentes sur Wikipédia, facilitant la mise en place de services multilingues sur les sites internet utilisant cette technologie.

DBpédia en français est désormais disponible dans le Web des données liées et ouvertes (Linked Open Data). Elle est réutilisée par des acteurs publics et privés ainsi que par les particuliers, par exemple Izipedia, zone47, Sépage, Vocal Apps, etc. Elle a permis la réalisation du site expérimental JocondeLab qui donne accès, en 14 langues, à plus de 300 000 notices illustrées extraites de la base Joconde. La mise en ligne de ce projet expérimental, porté par le ministère de la Culture et de la Communication, est prévue pour Janvier 2014.

Cette première étape de collaboration inter-institutionnelle a donc largement démontré son efficacité et sa fécondité en matière d'innovation. La convention Sémanticpédia, annuelle, est reconductible, mais n'intègre pour le moment, que l'administration centrale du ministère de la Culture et de la Communication. Ce dispositif doit être renforcé et permettre la dynamisation des projets de R&D soutenus par l'ensemble des services du ministère notamment ses services déconcentrés mais également ses établissements publics sous tutelle.

Favoriser une telle évolution aura un rôle incitatif en aidant l'ensemble des services du ministère de la Culture et de la Communication et ses établissements publics sous tutelle à prendre en considération les enjeux de la R&D appliquée au Web sémantique. Elle leur fournira un cadre de collaboration pour élaborer et affiner leurs projets. Elle permettra de mutualiser des moyens, des savoir-faire et des développements technologiques afin de développer des projets de R&D au plan national. Ce dispositif favorisera ainsi les collaborations inter-

institutionnelles sur l'ensemble du territoire.

Le dispositif devra répondre à trois missions fondamentales pour la R&D, notamment :

- La mise en place, aux plans national et international, d'une veille permanente sur la recherche, la normalisation et les développements relatifs aux données culturelles sur le Web et aux applications. Il s'agira aussi de rassembler au plan national des jeux de données, des cas d'usages et des scénarios motivants, pour identifier les opportunités et les attentes des différents acteurs.
- La transformation des verrous scientifiques identifiés dans le champ des données culturelles, en projets de R&D pour les sciences du numérique dans le domaine du Web sémantique.
- Le transfert et la valorisation des résultats depuis le monde de la recherche vers les acteurs de la culture pour une mise en œuvre effective. Ceci passe par des actions pour favoriser le déploiement d'innovations réussies par une institution pilote à travers l'ensemble des institutions. Cette démarche doit être soutenue par un plan de sensibilisation, de communication et d'accompagnement afin que tous les acteurs, quels que soient leurs profils et leurs tailles, tirent pleinement parti des résultats de la R&D.

ENJEUX STRATEGIQUES

Inclure l'ensemble des services du ministère de la Culture et de la Communication et des établissements publics sous tutelle dans le dispositif national de R&D structuré, consacré à l'application des technologies du Web sémantique aux données et usages culturels

Renforcer la collaboration entre les services, établissements et organismes culturels

RISQUES A NE PAS FAIRE

Laisser s'installer au sein du ministère de la Culture et de la Communication une démarche centralisatrice dans la stratégie de R&D appliquée au Web sémantique culturel

POINTS D'ATTENTION

Nécessité de mettre en place un plan d'accompagnement et de sensibilisation à l'attention des services du ministère de la Culture et de la Communication et de ses établissements publics afin de pouvoir pleinement tirer parti de ce dispositif

LIEN AVEC D'AUTRES ACTIONS

Action 9 : Mettre en place un programme de sensibilisation et de formation interne pour les services du ministère de la Culture et de la Communication et les établissements publics sous tutelle

LIVRABLES DU GROUPE DE TRAVAIL SPECIALISE

Proposer une évolution du cadre conventionnelle Sémanticpedia

Sensibiliser les acteurs culturels aux enjeux et aux principes de la R&D appliquée au Web sémantique culturel

ACTION 7 APPLIQUER LES TECHNOLOGIES 3.0 A LA DESCRIPTION DES EVENEMENTS CULTURELS

RESUME

De nombreux services du ministère de la Culture et de la Communication et établissements publics sous tutelle sont impliqués dans des événements culturels. Les données qu'ils produisent ou détiennent pour la diffusion et l'archivage de ces événements constituent une part essentielle du Web culturel. L'application des technologies 3.0 à ces données produira une forte valeur sur le plan des usages.

DESCRIPTION

De nombreux services du ministère de la Culture et de la Communication et établissements publics sous tutelle ont pour mission de créer des événements culturels. A l'échelle nationale, ces événements sont nombreux et traversent l'ensemble des domaines de la culture. Ils couvrent largement le territoire et concourent puissamment au dynamisme du tourisme culturel.

Concerts, festivals, représentations théâtrales, expositions, cours, ateliers, stages, conférences, etc, ces événements sont variés mais présentent des caractéristiques communes : ils rassemblent des personnes (auteurs, interprètes, acteurs, intervenants, etc.) qui sont bien souvent mises en relation avec des œuvres du patrimoine (plastiques, musicales, littéraires, etc.) éventuellement autour d'un thème, dans un ou plusieurs lieux, à une ou plusieurs dates.

Afin d'organiser et de promouvoir ces événements, les institutions culturelles produisent sur chacun d'entre eux une grande quantité de données. Pour créer et diffuser ces données, qui sont cruciales pendant toute la période de l'organisation de l'événement jusqu'à sa fin, chaque institution utilise des outils informatiques adaptés à sa taille et à son domaine d'activité. Les données produites en vue de l'archivage des événements sont souvent gérées à l'aide d'outils informatiques différents, ce qui implique des ressaisies.

L'utilisation des technologies du Web 3.0 pour la création et la diffusion de ces données présenterait de grands avantages. D'une part, les descriptions des événements culturels seraient plus homogènes d'un établissement à l'autre, et donc plus réutilisables. D'autre part, elles pourraient être judicieusement liées non seulement aux référentiels internes mais aussi à des référentiels externes du Web des données liées (potentiellement liés entre eux) pour désigner les personnes (ex. ISNI, VIAF, etc.), les œuvres (data.bnf, MusicBrainz, etc.), les lieux (Dbpédia, Geonames, etc.) ou les thèmes associés aux événements.

Les bénéfices de cette démarche sont nombreux tels :

- Favoriser l'émergence de nouveaux outils de recherche événementiels performants couvrant l'ensemble d'un territoire ou d'un domaine culturel ;
- Assurer l'intégration de ces données dans les graphes de connaissance des moteurs de recherche généralistes qui permettent, par exemple, d'obtenir instantanément toutes les dates et lieux de la tournée d'un artiste ;
- Faciliter l'accès multilingue et l'appropriation de la vie culturelle par un public international ;
- Développer la description des « traces » laissées par l'événement (enregistrements, images, affiches, etc.), à l'intérieur comme à l'extérieur des établissements producteurs des événements ;
- Fournir plus facilement aux publics des ressources numériques existantes en lien avec

l'événement avant, pendant et après.

ENJEUX STRATEGIQUES

Visibilité de la vie culturelle française dans les territoires comme au plan international

Meilleure connaissance de la vie culturelle à travers les événements

Promouvoir l'innovation économique

RISQUES A NE PAS FAIRE

Renforcer une logique de silo entre les acteurs organisateurs d'événements culturels et les territoires impactés

Priver des avantages d'un soutien national, les services du ministère de la Culture et de la Communication et les établissements publics sous tutelle les plus concernés par l'événementiel

POINTS D'ATTENTION

L'adoption de ces technologies doit être progressive et respecter les contraintes commerciales de vente et de promotion des événements

LIEN AVEC D'AUTRES ACTIONS

Action 2 : Développer un dispositif d'identification des auteurs de ressources culturelles

Action 3 : Créer et maintenir une interconnexion sémantique des grands référentiels culturels : le « Graphe Culture »

Action 4 : Expérimenter l'interconnexion inter-institutionnelle des données culturelles

Action 6 : Mettre en place un dispositif global de R&D destiné à soutenir l'innovation du Web 3.0 culturel

LIVRABLES DU GROUPE DE TRAVAIL SPECIALISE

Lancement d'une consultation publique sur la gestion d'événements impliquant le Web des données liées

Rédaction d'une note de cadrage pour un état de l'art, au plan international, sur les technologies 3.0 applicables à la gestion d'événements.

Rédaction d'un guide à l'attention des services, établissements et organismes culturels pour une politique de sémantisation des événements culturels

ACTION 8 POSITIONNER LE MINISTERE DE LA CULTURE ET DE LA COMMUNICATION EN TANT QU'EXPERT SUR LA TRACABILITE DES DONNEES NUMERIQUES

RESUME

Pour utiliser et archiver en toute confiance des données numériques, il est nécessaire de disposer d'informations sur leur traçabilité (origine, contexte de production etc.). Ces métadonnées de traçabilité font l'objet de travaux de normalisation au niveau international. Le ministère de la Culture et de la Communication, qui a la charge de conserver à très long terme les données patrimoniales, doit se positionner en tant qu'expert sur l'implémentation des normes de traçabilité.

DESCRIPTION

Le développement du Web de données, en impliquant la mise à disposition d'une quantité croissante de données issues d'acteurs variés, entraîne naturellement l'émergence de problématiques de sécurité, de traçabilité et de gestion temporelle qui sont communes avec le Web en général : pour réutiliser des données, il est nécessaire de savoir qui les a produites, dans quel contexte et avec quel degré d'expertise, à quel moment, si elles ont été mises à jour ou au contraire s'il existe un risque d'obsolescence, etc.

L'ensemble de ces informations qui permettent la traçabilité des données sur laquelle repose la confiance sont désignées par le terme d'« informations de provenance ».

Cette problématique était dès l'origine présente dans la représentation des différentes « briques » constitutives du Web sémantique (*semantic Web stack*), envisagées comme une clef de voûte nécessaire au bon fonctionnement de l'ensemble. En 2010, le *World Wide Web Consortium* (W3C) s'est lancé dans un chantier de normalisation voué à la traçabilité des données numériques qui a débouché sur un modèle de données connu sous le nom de « *PROV-O* » ou « *PROV Ontology* ». L'objectif de cette ontologie est de normaliser la création et la publication des informations de traçabilité dans le cadre du Web sémantique. L'existence systématique de telles informations permettra de déterminer des secteurs de confiance dans le Web de données et d'en assurer la gestion sur le long terme.

Cette problématique concerne au plus haut point les autorités publiques qui doivent assurer la pérennité et la transparence des processus de production et de publication des données. Il est probable qu'à l'instar des recommandations d'accessibilité du W3C (*Web Accessibility Initiative*) qui ont abouti en France au Référentiel Général d'Accessibilité pour les Administrations (RGAA), les travaux issus du groupe de travail « *PROV-O* » se déclineront au niveau des États sous la forme de référentiels nationaux.

La vocation patrimoniale du ministère de la Culture et de la Communication en fait l'un des acteurs désignés pour prendre en main cette problématique transverse à tous les domaines culturels : patrimoine, industries culturelles et création artistique contemporaine. Si les bibliothèques et les musées ont développé une expertise particulière dans l'expression de métadonnées destinées à la gestion à long terme de leurs catalogues et inventaires, dans les archives cette problématique est encore plus centrale pour la fonction support qu'est l'archivage des documents et données numériques. La mission « traditionnelle » des archivistes, qui est d'interroger la qualité des données afin de pouvoir attester qu'ils prennent en charge des « objets » dignes de confiance, revêt un intérêt encore plus évident face aux

données numériques qu'il est très facile de falsifier, d'altérer intentionnellement ou non, ou encore de perdre en raison d'une obsolescence technologique toujours plus rapide. Il est indispensable d'intervenir très en amont de manière à pouvoir définir et mettre en œuvre le cycle de vie des données numériques ainsi qu'une stratégie d'archivage.

La pérennisation ne porte pas uniquement sur des questions d'organisation et de stratégie, de formats de fichiers ou des supports sur lesquels enregistrer l'information. Il s'agit aussi de décrire les documents et données numériques pour assurer leur traçabilité et connaître le contexte de leur création ainsi que l'historique des traitements et interventions opérés sur ces documents et données.

Dans le cadre de son partenariat avec Inria, qui est historiquement très impliqué dans les actions de normalisation du W3C, le ministère de la Culture et de la Communication peut se positionner en collaboration d'expertise sur ce sujet. Cette implication peut prendre la forme d'une cellule d'experts du secteur culturel (Sous-direction des systèmes d'information, Service interministériel des Archives de France, etc.) aux côtés d'experts Inria. Le ministère de la Culture et de la Communication sera alors à même de jouer un rôle majeur pour la transposition nationale des recommandations du W3C lorsque le processus de normalisation sera entré en phase d'implémentation nationale. Cette démarche ne devra pas se limiter à l'implémentation de la norme « *PROV-O* », mais concerner tout processus normatif du W3C lié à la traçabilité (*WebID* par exemple).

ENJEUX STRATEGIQUES

Faciliter la réutilisation et l'exploitation des données en garantissant aux réutilisateurs que les données fournies sont fiables et donc traçables

Assurer une gestion de long terme des données produites et échangées dans le cadre du Web sémantique

RISQUES A NE PAS FAIRE

Risque de divergence entre les pratiques métier de l'archivage numérique et les pratiques numériques qui seraient déconnectées des besoins à long terme

Risque que l'archivage des données du Web sémantique institutionnel et culturel français soit complexifié, voire rendu impossible par l'absence de traçabilité

POINTS D'ATTENTION

Pouvoir marquer dès la conception des systèmes d'information la qualité d'une donnée susceptible d'être réutilisée

Analyser les données à travers leur processus métier et s'assurer qu'un contrôle préalable de ce processus existe

LIEN AVEC D'AUTRES ACTIONS

Action 1 : Mettre en place un dispositif d'identifiants pérennes pour les ressources culturelles

Action 2 : Développer un dispositif d'identification des auteurs de ressources culturelles

Action 9 : Mettre en place un programme de sensibilisation et de formation interne pour le ministère de la Culture et de la Communication

LIVRABLES DU GROUPE DE TRAVAIL SPECIALISE

Production d'une note de cadrage relative à l'état de l'art sur la normalisation du W3C en matière de traçabilité des données

Préfiguration d'une cellule de veille permanente sur la traçabilité des données

ACTION 9 METTRE EN PLACE UN PROGRAMME DE SENSIBILISATION ET DE FORMATION INTERNE POUR LE MINISTÈRE DE LA CULTURE ET DE LA COMMUNICATION

RESUME

Pour développer au sein du ministère de la Culture et de la Communication et de ses établissements publics une véritable culture commune du changement et de l'innovation autour du Web sémantique, il est nécessaire de mettre en place un plan de formation ambitieux, interdisciplinaire et inter-institutionnel.

DESCRIPTION

Les usages, les technologies mais aussi les enjeux du Web 3.0 sont encore mal connus au sein des institutions culturelles. Comme pour grandes étapes précédentes de l'histoire du Web – le Web des documents et le Web social (ou Web 2.0) – le Web 3.0 entraîne une évolution des pratiques et des outils qui nécessite de sensibiliser les décideurs et d'accroître la compétence des agents par des actions de formation. Mais à la différence des étapes précédentes, le principe de l'interconnexion des gisements de données induit une dimension inter-institutionnelle et interdisciplinaire qui doit impérativement être prise en compte dans les plans d'accompagnement du changement.

Pour développer une approche cohérente, faciliter l'émergence de projets traversant les disciplines et les organisations, partager les expériences, les pratiques et les outils, il est nécessaire de développer au sein du ministère et de ses établissements une véritable culture commune du changement et de l'innovation et de faire monter en compétence de façon concertée l'ensemble des agents impliqués dans des projets Web 3.0.

C'est pourquoi un plan de formation doit être rapidement défini et mis en œuvre. Il comprendra des actions de communication interne, d'information et de sensibilisation en direction des cadres, responsables scientifiques et instances décisionnaires, afin de les amener à comprendre les enjeux, et les intégrer dans leur stratégie de développement. Les services des ressources humaines du ministère de la Culture et de la Communication et des établissements publics sous tutelle ainsi que la sous-direction des systèmes d'information seront pour ce faire sollicités.

Le plan de formation devra également permettre la montée en compétences des différents acteurs des projets, qu'il s'agisse des spécialistes de domaines ou des spécialistes de l'information et de la documentation, impliqués dans le choix des contenus, la définition de nouveaux usages, la modélisation conceptuelle, l'identification, la consolidation et la gestion des référentiels. Les chefs de projet en charge de la maîtrise d'ouvrage et de la maîtrise d'œuvre de projets des systèmes d'information (SI) devront être en mesure de prendre en compte la spécificité des projets Web 3.0 lors de la définition des projets, la rédaction des cahiers des charges et des exigences fonctionnelles et techniques. Quant aux responsables informatiques, ils devront être sensibilisés aux problématiques de changement d'architecture et de disponibilité des services d'exposition induites par l'émergence du Web 3.0 dans les SI et s'assurer que ces nouvelles contraintes seront prises en compte et maîtrisées par les services d'exploitation.

ENJEUX STRATEGIQUES

Mise en place d'un programme de sensibilisation et de formation pour le ministère de la Culture et de la Communication

Mutualisation de certaines sessions de formation entre les différentes entités du ministère de la Culture et de la Communication

RISQUES A NE PAS FAIRE

Isolement des services et établissements du ministère de la Culture et de la Communication n'ayant pas les ressources nécessaires à la mise en place d'un plan de sensibilisation et de formation

POINTS D'ATTENTION

Développer une culture du changement nécessite une adhésion de l'ensemble des acteurs et un soutien des structures décisionnelles

LIEN AVEC D'AUTRES ACTIONS

Action 6 : Mettre en place un programme de R&D destiné à soutenir l'innovation du Web 3.0 culturel

LIVRABLES DU GROUPE DE TRAVAIL SPECIALISE

Pré-programme pour un portefeuille de formation impliquant l'ensemble des services et établissements du ministère de la Culture et de la Communication

Pré-programme de communication interne à destination des agents du ministère de la Culture et de la Communication

ANNEXE

FICHE DE PRESENTATION

Sémanticpédia

Sémanticpédia est une **plate-forme de collaboration** entre le ministère de la Culture et de la Communication, Inria et Wikimedia France pour réaliser des programmes de recherche et développement appliqués à des corpus ou des projets collaboratifs culturels, utilisant des données extraites des projets de Wikimedia.

La convention de partenariat a été signée le 19 novembre 2012 avec Aurélie Filippetti, ministre de la Culture et de la Communication, lors de la journée de lancement du projet DBpédia en français, qui constitue le premier projet réalisé dans ce cadre.

Ce partenariat, qui s'inscrit dans la démarche d'ensemble d'une politique culturelle de l'accès, constitue d'une part un enjeu pour la facilitation de l'accès du public à des données culturelles en français, et, d'autre part, un enjeu pour le développement d'une politique du multilinguisme, les données liées permettant de favoriser l'offre de contenus multilingues et d'apporter une contribution significative à la diversité linguistique.

Rôle et apports des partenaires

- **Le ministère de la Culture et de la Communication** apporte l'expertise de plusieurs de ses départements sur des contenus ou projets culturels en lien avec un réseau national d'établissements culturels. Il est moteur pour la définition de projets.
 - Département des programmes numériques* : expertise sur les stratégies de Web sémantique et de développement des formats RDF ; premier ré-utilisateur du projet avec le site Histoire des Arts et la preuve de concept HDA-Lab ;
 - Sous-direction des systèmes d'information* : expertise technique sur les questions d'ouverture et d'inter-opérabilité des données, sur la réutilisation et l'intégration des technologies sémantiques dans les projets du MCC et notamment Culture.fr ;
 - Délégation générale à la langue française et aux langues de France* : expertise sur les enjeux liés à la présence du français et au développement du multilinguisme sur l'internet suivi des aspects purement linguistiques ; prise en charge de la communication institutionnelle ; suivi global du projet, Pilotage du programme « Sémantisation » du MCC.
- **Wikimédia France** apporte son expertise sur les corpus, les outils, les données, les processus éditoriaux et la plate-forme wiki des projets wikimédiens. Elle affirme son objectif de soutien du projet « DBpédia.fr » portant sur les contenus sémantisés de Wikipédia. Elle favorise également la création d'une communauté de dialogue et d'échange entre le MCC, Inria et les contributeurs des projets Wikimédia afin d'optimiser le développement des projets réalisés dans le cadre de la plate-forme « Sémanticpédia » afin que ceux-ci soient réellement en phase avec la réalité des projets Wikimédia et de leurs contributeurs et ainsi extraire avec une qualité optimale les contenus sémantisés des projets Wikimédia.
- **Inria**, à travers l'équipe Wimmics, développe des outils expérimentaux mobilisant des données et services culturels du MCC, d'une part, et des données et services des projets wikimédiens, d'autre part. En plus du soutien scientifique et technique déjà

apporté par Inria à travers l'équipe de recherche Wimmics durant la phase de développement et l'hébergement initial, dans le cadre de la plate-forme « Sémanticpédia » et de ses futurs projets, l'institut pourra apporter les compétences de plusieurs de ses équipes de recherche et services de soutien en fonction des évolutions des projets de la plate-forme « Sémanticpédia ». Enfin, de par sa position notamment au sein du W3C, de la communauté scientifique du Web de données et du Web sémantique, et de la communauté du projet DBpedia.org, Inria pourra effectuer des activités de liaisons pour les projets de la plate-forme « Sémanticpédia ».

Projets réalisés dans le cadre de la convention Sémanticpédia

- **DBpédia en français** : DBpédia en français est le chapitre francophone de DBpédia, il s'inscrit dans l'effort d'internationalisation de DBpédia dont le but est de maintenir des données structurées extraites de différents chapitres de Wikipedia. Développé par Inria avec le concours de ses partenaires, le site est actuellement hébergé par l'association Wikimedia France.
- **Wiktionnaire sémantisé** : Le Wiktionnaire est un des projets de la Wikimedia Foundation visant à construire un dictionnaire multilingue libre et écrit collaborativement. Inria, avec l'appui de ses partenaires, a produit en 2013 une extraction complète du Wiktionnaire en format RDF, à partir de l'entrée en français.
- **HDA-Lab** : Le portail Histoire des arts a pour vocation de mettre à la disposition de publics prescripteurs — via un site Internet — une sélection de ressources culturelles et éducatives riches en lien avec le programme officiel de l'enseignement de l'Histoire des arts. Le volet HDA-Lab de ce projet s'inscrit dans le cadre d'une collaboration entre l'Institut de Recherche et d'Innovation (IRI) et le Ministère de la Culture et de la Communication. Ce projet de recherche et développement prévoit la réalisation d'une preuve de concept destinée à montrer le potentiel heuristique du « *tagging* » sémantique.
- **Muséosphère** : Le projet a été conçu par Mara Dumitru en 2011, pendant ses études en informatique à l'Université américaine de Paris. Tout au long du cours en Web sémantique, la passion de Mara pour l'art et pour la technologie lui a inspiré d'initier et de développer une application en Java, dénommée à l'époque Museum Recommender. En 2013, le MCC a offert à Mara Dumitru l'opportunité d'implémenter son projet dans le cadre de la plate-forme collaborative Sémanticpédia. Ainsi, le projet d'expérimentation prenait vie grâce à SÉPAGE, société spécialisée dans les technologies du Web sémantique.
- **JocondeLab** : Le site internet JocondeLab est le fruit d'une expérimentation menée par le ministère de la Culture et de la Communication. Il vise à démontrer les possibilités du Web sémantique ou « Web 3.0 » et de la mise en relation de données culturelles « liées », tant en termes de multilinguisme, que d'ergonomie et d'interactivité. L'expérimentation vise notamment à faciliter l'accès aux ressources culturelles françaises auprès de nouveaux publics et apporte une contribution significative à la diversité linguistique en développant l'offre multilingue du ministère de la Culture et de la communication, sans pour autant qu'il soit nécessaire de traduire les données puisque celles-ci trouvent leur équivalent grâce à Wikipédia.

Reconduction et évolution de la convention

La convention, signée le 19 novembre 2012, stipule qu'elle est conclue pour une durée d'un an et qu'elle se renouvelle de façon tacite chaque année, sauf si l'une des parties s'y oppose. La convention a donc été reconduite jusqu'au 19 novembre 2014.

Il est notamment précisé à ce sujet que : « Un bilan de l'expérimentation établi chaque année à la date d'anniversaire de la convention par Wikimedia France aidera les Parties à se décider des suites à donner à leur collaboration. Les modalités de cette évaluation seront à définir entre les Parties en fonction des outils à leur disposition lorsque les applications développées dans le cadre de la plate-forme « Sémanticpédia » seront utilisables, et des premiers résultats qui serviront de base de référence. L'évaluation pourra notamment porter sur le nombre de requêtes d'utilisateurs sur le service « DBpédia.fr », mais également sur le nombre de ré-utilisateurs officiels (ré-utilisateurs industriels, culturels ou institutionnels) ainsi qu'un suivi des retombées du projet en termes de communication. ».

FEUILLE DE ROUTE STRATEGIQUE
Métadonnées culturelles et transition Web 3.0



Ministère de la Culture et de la Communication
Secrétariat général
Département des programmes numériques
182, rue Saint-Honoré, 75033 Paris Cedex 01