



SENSIBILISATION ET FORMATION DES ACTEURS CULTURELS

RÉFÉRENTIEL DE FORMATION

WEB SÉMANTIQUE,

WEB DES DONNÉES

Ministère de la Culture et de la Communication
Stratégie « Métadonnées culturelles et transition Web 3.0 »



Publié en décembre 2015

Ce document est mis à disposition sous licence CC BY-SA 3.0 FR
(<https://creativecommons.org/licenses/by-sa/3.0/fr/>)

INTRODUCTION

Le « Web 3.0 », dit aussi « Web sémantique » ou « Web des données », repose sur une extension des normes fondamentales du web. Cette évolution technologique, la plus importante depuis la création du web au début des années 1990, est en train de transformer le web en une gigantesque base de connaissance. Elle permet, entre autres, la réalisation d'outils recherche d'information, de navigation, de visualisation, de traduction et à terme de raisonnements automatiques, hors de portée des moteurs de recherche classiques.

Elle est cependant quasiment inconnue du grand public (à la différence du Web 2.0, par exemple) car elle impacte en premier lieu l'écosystème profond du web, ce que l'on appelle « le web invisible », c'est-à-dire les gisements d'informations contenus dans les bases de données sous-jacentes au web. A terme, le Web 3.0 impactera fortement le champ culturel, tant sur le plan de l'organisation des ressources numériques que sur celui des usages, des outils et des infrastructures.

Dans cette perspective, la feuille de route ministérielle « Métadonnées culturelles et transition Web 3.0 »¹ propose une vingtaine d'actions-clés à mener à court-moyen terme. L'action n° 9 de la feuille de route a notamment pour objectif de mettre en place un programme de formation transverse destiné à l'ensemble des personnels potentiellement ou directement acteurs dans ces évolutions au sein du ministère de la Culture et de la Communication (MCC) et de ses établissements sous tutelle. Le développement des compétences apparaît en effet d'autant plus décisif que les spécificités du Web 3.0 sont encore trop peu connues, et encore moins techniquement maîtrisées, au sein des institutions culturelles.

Pour répondre à ce besoin, un groupe d'experts, piloté par le Département du recrutement, de la mobilité et de la formation du MCC, a élaboré un référentiel de formation consacré aux technologies orientées Web 3.0. Ce document est destiné à servir de base à la construction de l'offre de formation du secrétariat général, ouverte à l'ensemble des structures du ministère (administration centrale, directions régionales des affaires culturelles, établissements publics, services à compétence nationale, etc.). L'enjeu de ce programme de formation transverse est :

- de faire monter en compétence, de façon homogène, l'ensemble des agents impliqués dans des projets orientés Web 3.0,
- de développer au sein du ministère et de ses établissements une culture commune du changement et de l'innovation,
- de faciliter l'émergence de projets traversant les disciplines et les organisations.

Le référentiel est composé de quatre modules. Le premier est à la fois un stage d'initiation aux enjeux du web 3.0 et la première étape d'un parcours constitué ensuite de trois modules spécialisés s'adressant plus spécifiquement à certains acteurs dont les métiers se trouvent au cœur de ces évolutions :

- Les chefs de projet Maîtrise d'ouvrage (MOA)
- Les chefs de projet Maîtrise d'œuvre
- Les professionnels de l'information-documentation

Dans la mesure où ce référentiel peut être utile dans d'autres contextes (ministères, organismes publics ou privés, etc.), celui-ci est publié sous licence CC By SA afin d'en garantir la plus grande réutilisabilité.

1. http://www.culturecommunication.gouv.fr/var/culture/storage/pub/feuille_de_route_metadonnees_culturelles_et_transition_web_3_0_janvier_2014/index.htm#/38

Membres du groupe de travail

Experts du ministère de la Culture et de la Communication :

Président du groupe de travail : *Fabien Grimaud (SG)*

Naïma Alim (DGCA), Katell Briatte (DGP), Claire Choppe (SG), Jean-Pierre Defrance (DGP), Madé Haguma (SG), Laurence Korenian (SG), Annie Lelandais (DGMIC), Odile Levassor (SG), Eric Rouard(SG), Bertrand Sajus (SG), Anna Svenbro (DGMIC), Brigitte Tran (DGLFLF)

Experts externes consultés :

Jean Charlet (Inserm), Nicolas Chauvat (Logilab), Sylvie Dalbin (ATD, ADBS), Jean-Pierre Déclé (www.jpdeclé.com), Jean Delahousse (jean-delahousse.net), Thomas Francart (Sparna), Hélène Rabault (Lexis Nexis)

Nous tenons à remercier l'ensemble des experts externes qui ont participé à ce référentiel à titre bénévole et particulièrement M. Jean Delahousse pour sa précieuse contribution tout au long de nos travaux.

SOMMAIRE

Initiation aux enjeux du Web 3.0	p. 4
Chefs de projet Maîtrise d'ouvrage (MOA)	p. 6
Chefs de projet Maîtrise d'œuvre (MOE)	p. 9
Professionnels de l'information-documentation	p. 12

Web 3.0 : Initiation aux enjeux

CONTEXTE

Le Web 3.0 entraîne une évolution des pratiques et des outils qui nécessite de sensibiliser les décideurs et d'accroître la compétence des différents acteurs par des actions de formation. Pour développer une culture commune du changement et de l'innovation autour du Web sémantique, le choix a été fait de mettre en place un plan de formation interdisciplinaire et inter-institutionnel, qui intègre l'ensemble des acteurs de cette évolution et permette la montée en compétences des différents acteurs des projets du Web 3.0.

Le Web 3.0 est un moment crucial de l'évolution générale du Web mais ses enjeux sont très peu connus du grand public. A la différence du Web dit «1.0» des années 1990, et du Web «2.0» des années 2000, il repose sur des mécanismes et des notions (telles que les «données liées», par exemple) beaucoup moins intuitifs que les notions de pages HTML ou encore de «réseau social». Cela tient en grande partie au fait que le Web 3.0 concerne d'abord les «données» et surtout leur interconnexion, et non pas directement les sources d'informations et les documents tels qu'ils se présentent à l'internaute via les navigateurs. Il requiert donc un effort de vulgarisation tout particulier.

C'est pourquoi il apparaît nécessaire d'offrir une formation courte, ouverte largement, permettant aux participants de comprendre rapidement les notions et les enjeux fondamentaux du Web 3.0.

Ce stage doit également fournir les prérequis aux trois cibles métiers du plan de formation Web 3.0 : professionnels de l'information-documentation, chefs de projets MOA, chefs de projets MOE.

PUBLIC CIBLE

- Toute personne qui souhaite mieux comprendre les grandes évolutions et orientations technologiques du Web et les enjeux pour la structure.
- La priorité doit néanmoins être donnée aux personnes appartenant aux trois cibles métiers identifiées dont cette formation est la première pierre du parcours

PREREQUIS EVENTUELS

Aucun

OBJECTIFS GENERAUX DE LA FORMATION

Faire comprendre les notions clés du Web 3.0 et ses enjeux, notamment pour le secteur culturel

Cet objectif peut se décliner en :

- comprendre en quoi le Web 3.0 est déjà une réalité dans le Web actuel
- comprendre les notions techniques sur lesquelles reposent le Web 3.0
- comprendre les enjeux particuliers dans le domaine culturel

OBJECTIFS PEDAGOGIQUES

A l'issue de ce stage, les participants doivent être capables de :

- savoir situer le Web 3.0 dans l'évolution globale du Web
- identifier les notions clés du Web 3.0
- appréhender les enjeux du Web 3.0 notamment dans le domaine culturel
- appréhender les orientations stratégiques pour la structure (pour le MCC)
- savoir utiliser quelques applications Web 3.0

ELEMENTS DU PROGRAMME

Thèmes/contenus	Durée estimée
<p>Découverte des quelques applications Web 3.0</p> <ul style="list-style-type: none"> ▪ Data.bnf.fr ▪ Google et le web sémantique: Knowledge Graph et l'utilisation de schema.org ▪ Zemanta ▪ HDA-Lab et JocondeLab 	1 h
<p>Le Web 3.0 dans l'évolution globale du Web</p> <ul style="list-style-type: none"> ▪ Rappel des grandes étapes historiques de la construction du Web : continuité internet – Web des documents – Web social – Web des données ▪ La mise en évidence de l'émergence du Web 3.0 dans cette évolution globale 	30 mn
<p>Identification des notions clés du Web 3.0</p> <ul style="list-style-type: none"> ▪ La notion de donnée ▪ La gestion des identifiants: URI et sur la différence entre URL et URI ▪ Le triplet et le RDF ▪ Le langage d'interrogation (SPARQL) ▪ Les ontologies (OWL) ▪ Les usages des moteurs d'inférence 	1h30
<p>Les ontologies essentielles</p> <ul style="list-style-type: none"> ▪ Dublin Core, SKOS, FOAF, FRBR ▪ Compréhension du processus d'utilisation d'une ontologie ▪ L'interopérabilité inter institutions par les ontologies 	1 h
<p>Les vocabulaires contrôlés essentiels (dans le monde culturel)</p> <ul style="list-style-type: none"> ▪ Rameau, Library of Congress Subject Headings, Getty, Geonames, VIAF... ▪ Réutilisation de vocabulaires contrôlés et alignement ▪ L'interopérabilité inter institutions par les vocabulaires 	1 h
<p>Appréhension des enjeux du Web 3.0</p> <ul style="list-style-type: none"> ▪ Les enjeux du Web des données liées ▪ La spécificité du Web sémantique par rapport au Web des données liées ▪ La notion d'écosystème 3.0 ▪ Un focus sur quelques carrefours du Web 3.0 culturel : Dbpedia, data.bnf.fr, Getty LOD, Europeana... ▪ Futur proche : les conséquences du web des données et les changements induits 	1h30

Web 3.0 : Chefs de projet Maîtrise d'ouvrage (MOA)

CONTEXTE

Le Web 3.0 entraîne une évolution des pratiques et des outils qui nécessite de sensibiliser les décideurs et d'accroître la compétence des différents acteurs par des actions de formation. Pour développer une culture commune du changement et de l'innovation autour du Web sémantique, il apparaît nécessaire de mettre en place un plan de formation interdisciplinaire et inter-institutionnel, qui intègre l'ensemble des acteurs de cette évolution et permette la montée en compétences des différents acteurs des projets du Web 3.0.

Parmi ces acteurs, les chefs de projet MOA devront être en mesure de prendre en compte la spécificité des projets Web 3.0 lors de la définition des projets.

Ils doivent tout particulièrement être en capacité de comprendre et d'anticiper cette évolution, d'en intégrer les enjeux en tant que porteurs d'innovation dans ce domaine et auprès des commanditaires et des utilisateurs.

PUBLIC CIBLE

Chefs de projet maîtrise d'ouvrage

PREREQUIS EVENTUELS

Connaissance des bases et enjeux du web sémantique et du web des données

OBJECTIFS GENERAUX DE LA FORMATION

Être en capacité d'assumer le rôle d'interface entre les utilisateurs, les commanditaires, la MOE et les prestataires sur les problématiques du web 3.0 :

- Comprendre les technologies du web sémantique (concepts, langages)
- Comprendre les notions de données ouvertes (open data), de web des données liées (linked open data) et les contraintes associées (pérennité, fiabilité, réutilisabilité)
- Connaître et savoir utiliser certains outils; outils de validation, éditeurs d'ontologie, outils de sémantisation de données, etc.
- Savoir évaluer la valeur ajoutée des technologies et des projets sémantiques pour conseiller les commanditaires et les futurs utilisateurs d'un projet
- Savoir formaliser les exigences et les spécifications fonctionnelles d'un projet qui utilise les technologies du Web Sémantique
- Savoir intégrer la dimension sémantique dans les problématiques de reprise de données et définir un schéma de conversion des données au format sémantique

OBJECTIFS PEDAGOGIQUES

- Comprendre les apports du « web des données » par rapport au « web des documents »
- Connaître les différents types d'utilisation des technologies du web sémantique
 - publication de données
 - alignement de données
 - agrégation de données
 - recherche sémantique, etc.
- Avoir une connaissance générale de l'architecture du web sémantique
 - briques technologiques, normes et standards
 - ontologies les plus communément réutilisées
 - connaître les différents modes de sémantisation des données
 - microformats
 - RDF
 - ontologies
- identifier les données sémantisables (données structurées, référentiels)
- spécifier des interfaces utilisateurs exploitant les possibilités du Web Sémantique
- comprendre et anticiper les enjeux juridiques liés à l'ouverture des données

ELEMENTS DU PROGRAMME

Thèmes/contenus	Durée estimée
Concepts généraux du web sémantique dans le cadre de la gestion de projets <ul style="list-style-type: none"> ▪ Notions de bases ▪ Périmètre du web sémantiques ▪ Les enjeux 	30 mn
Analyse d'applications clefs du Web sémantique <ul style="list-style-type: none"> ▪ Analyse d'exemples d'applications <i>culturelles</i> ▪ Identification des différents types d'applications ▪ Les objectifs en termes d'usage et les potentialités ▪ Identification des apports du web sémantique ▪ Les limites, les risques et les contraintes liés à chaque type d'application 	2h
Méthodologie de projets web sémantique <ul style="list-style-type: none"> ▪ Les spécificités de ce type de projet ▪ Typologie des acteurs et leurs périmètres d'intervention ▪ Le principe du calcul du retour sur investissement sur ce type de projet ▪ <i>L'articulation avec la méthodologie de gestion de projets du MCC</i> 	1h
Les URI <ul style="list-style-type: none"> ▪ Négociation de contenu : Différence entre URL et URI ▪ Une ressource plusieurs représentations ▪ Identifiants pérennes ▪ Règles de nomage des URIs pour les différents types de ressources ▪ Redirection des URIs 	2h
Présentation détaillée des standards et des langages <ul style="list-style-type: none"> ▪ RDF, RDFa ▪ SPARQL ▪ RDFS, OWL 	3h
Conception d'une ontologie <ul style="list-style-type: none"> ▪ OWL, concevoir une ontologie ▪ Réutilisation d'ontologies et vocabulaires contrôlés existants ▪ Méthodologie de construction à partir d'un corpus culturel ▪ Utilisation d'un éditeur d'ontologies ▪ Outils de recherche d'ontologies « LOV » ▪ Ontologies couramment réutilisées: Dublin Core, Foaf, Skos, FRBR,Org, Music ontology, CIDOC CRM) 	9h
Migration / Publication des données dans des formats web sémantique <ul style="list-style-type: none"> ▪ Migration/publication de bases de données relationnelle ▪ Scénarios de sémantisation de données ▪ Outils de conversion/publication (sélection limitée d'outils) ▪ Cas pratique 	6h
Enrichissement sémantique de contenus <ul style="list-style-type: none"> ▪ Classification automatique de contenus ▪ Utilisation du text mining pour enrichir l'annotation sémantique du contenu ▪ Utilisation du text mining pour formaliser des connaissances 	1h30

Interfaces <ul style="list-style-type: none"> ▪ Exploitation du web sémantique pour la construction d'interfaces efficaces ▪ Recherche par facettes (interface et ergonomie) ▪ Fiche de connaissance/ « wikipédisation » ▪ Vues multiples sur les données (temporelle, spatiale, thématique...) ▪ Interface utilisateur 	2h
Alignements <ul style="list-style-type: none"> ▪ Aligner des ontologies ▪ Aligner des vocabulaires contrôlés; thésaurus, autorités, taxonomies ▪ Méthodes et outils pour l'alignement ▪ Publier des alignements ▪ Intéropérabilité sémantique 	2h
Écosystème, cartographie des graphes de connaissance <ul style="list-style-type: none"> ▪ Les référentiels et bases de connaissances communément réutilisées: dbpedia, FreeBase, Yago, Geonames, Subject Headings de la Library of Congress, Vial, IMDB, MusikBrainz, Rameau, Thésaurus du GETTY...) 	2h
Méthodes de mise à disposition des données sur le web des données <ul style="list-style-type: none"> ▪ Test et validation des données à publier ▪ Les différentes formes de distribution des données : fichiers, URIs, API http, service SPARQL, html/microdatas ▪ Quelles sérialisations des données RDF proposer ▪ Créer un portail open data ou utiliser un portail existant 	2h
Traçabilité et Sécurité juridique <ul style="list-style-type: none"> ▪ Prendre en compte les problématiques de traçabilité des données dans la définition d'un projet ▪ L'ontologie PROV-O pour la gestion de la traçabilité ▪ Vigilance sur la sécurité juridique (CNIL, etc.) 	1h
Cycle de vie des données / Versionning <ul style="list-style-type: none"> ▪ Evolution temporelle des données 	30 min
Les outils pour les MOA <ul style="list-style-type: none"> ▪ Valideurs <ul style="list-style-type: none"> - RDF : valider la qualité d'une sérialisation RDF - Valideur SKOS 	1h
Les acteurs / offres <ul style="list-style-type: none"> ▪ Cartographie des outils et services pour un projet qui utilise des technologies du web sémantique ▪ Fournisseurs 	30min
Veille <ul style="list-style-type: none"> ▪ Sites, listes de diffusion, blogs à connaître 	1h
Retour d'expérience sur un projet type (Projet à définir avec le ministère)	2h
Total	6 jours

Web 3.0 : Chefs de projet Maîtrise d'œuvre (MOE)

CONTEXTE

Le Web 3.0 entraîne une évolution des pratiques et des outils qui nécessite de sensibiliser les décideurs et d'accroître la compétence des différents acteurs par des actions de formation. Pour développer une culture commune du changement et de l'innovation autour du Web sémantique, il apparaît nécessaire de mettre en place un plan de formation interdisciplinaire et inter-institutionnel, qui intègre l'ensemble des acteurs de cette évolution et permette la montée en compétences des différents acteurs des projets du Web 3.0.

Parmi ces acteurs, les chefs de projets MOE doivent particulièrement être en capacité d'intégrer dans le projet, en cohérence avec les objectifs stratégiques de celui-ci, l'ensemble des potentiels du web sémantique et d'en préparer la mise en production et l'exploitation.

Ils doivent être en capacité de faciliter l'intégration des technologies «web 3.0» dans les projets informatiques dès que cela est pertinent, en accord avec le(s) schéma(s) d'urbanisation des systèmes d'information et de s'assurer de la pérennité des projets mis en œuvre.

PUBLIC CIBLE

Chef de projet Maîtrise d'œuvre

PREREQUIS EVENTUELS

- Connaissances de base sur les concepts et les standards du Web sémantique
- Connaissance des principales techniques de description des ressources et de publication

OBJECTIFS GENERAUX DE LA FORMATION

- Acquérir une vue d'ensemble concrète de la chaîne de mise en œuvre; description, création, publication de données sémantiques et utilisation au sein d'applications, outils disponibles et particulièrement représentatifs
- Connaître les grands choix d'architecture applicative : contraintes de cohérence liées, les avantages et les inconvénients.
- Connaître les principales problématiques d'une plate-forme de collecte et de diffusion de données : transformation, enrichissement, indexation, mise à l'échelle et mise à jour

OBJECTIFS PEDAGOGIQUES

- Connaître une gamme limitée d'outils sélectionnés comme représentatifs de l'état de l'art en étant capable d'en appréhender l'ensemble du fonctionnement (via des exercices pratiques notamment)
- Être capable d'identifier les principales techniques de description de ressources et de publication de données
- Être capable d'utiliser des modèles de description sémantique pour créer des applications exploitant des sources de données :
 - Savoir créer et lire un modèle RDF
 - Savoir exposer des données en tant que RDF
 - Être capable d'agréger des informations issues de différentes sources
 - Savoir utiliser des méthodes d'alignement des informations
- Connaître les principales options en matière d'utilisation et de publication de services sémantiques
 - Connaître les différents composants d'une application
 - Connaître le rôle des différents types d'outils : les apports, les contraintes liées.
En identifier les limites et les solutions en réponse (volumétrie, performances, mises à jour, etc)
 - Savoir comment exposer un service sémantique
- Comprendre les apports et contraintes des mécanismes de raisonnement et d'inférence

ELEMENTS DU PROGRAMME

Thèmes/contenus	Durée estimée
<p>Les applications : présentation générale</p> <ul style="list-style-type: none"> ▪ Types d'applications : Mashups, Web Mining, recherche, Wikis, Web Services ▪ Analyse d'exemples d'applications : grand public, d'entreprise et scientifiques ▪ Les différents composants d'une application (comparaison avec une architecture web classique) ▪ Les différents moteurs de recherche orientés Web 3.0 <p><i>Travaux pratiques</i> <i>Démonstration à partir d'exemple du Web sémantique.</i></p>	2h
<p>URIs</p> <ul style="list-style-type: none"> ▪ Négociation de contenu : Différence entre URL et URI ▪ Une ressource plusieurs représentations ▪ Identifiants pérennes ▪ Règles de nommage des URIs pour les différents types de ressources ▪ Redirection des URIs 	1h
<p>Format de description des données et ontologies</p> <ul style="list-style-type: none"> ▪ Les principales techniques de description de ressources et les outils de publication des données ▪ Données liées et modèles RDF : Linked data, RDF, RDFS, OWL, SPARQL, RDFa, GRDDL. (présentation et exemples) ▪ Les formats de sérialisation XML, N3, Turtle (exemples) ▪ Le stockage des triples RDF (triples stores) ▪ Les langages de règles (SWRL) 	3h
<p>Ontologies</p> <ul style="list-style-type: none"> ▪ RDFS / OWL ▪ Ontologies modulaires ▪ Présentation des logiciels de gestion d'ontologie ▪ L'exploitation des données de l'inférence 	3h
<p>Créer une ontologie :</p> <ul style="list-style-type: none"> ▪ Définir les classes qui seront utilisées dans l'ontologie ▪ Arranger les classes en une hiérarchie taxinomique ▪ Définir les attributs et décrire les valeurs autorisées pour ces attributs. ▪ Renseigner les valeurs pour les attributs des instances 	2h
<p>Créer un jeu de données RDF et l'utiliser</p> <ul style="list-style-type: none"> ▪ Déclarer un modèle ▪ Lire et écrire du RDF ▪ Contrôler la définition des préfixes ▪ Exécuter des requêtes dans un modèle 	2h
<p>Alignement d'ontologies et vocabulaires contrôlés</p> <ul style="list-style-type: none"> ▪ L'alignement pour répondre à des besoins d'interopérabilité ▪ Aligner les ontologies avec OWL et SWRL, RDFS ▪ Aligner des vocabulaires contrôlés; thésaurus, autorités, référentiels 	30 min
<p>Accéder à et créer un endpoint SPARQL</p> <ul style="list-style-type: none"> ▪ Le langage SPARQL ▪ Graphe nommé ▪ Créer un endpoint SPARQL <p><i>Travaux pratiques : utilisation d'une API (PHP ou Java) pour parser des descriptions RDF et créer un endpoint SPARQL.</i></p>	2h30

Le framework de développement <ul style="list-style-type: none"> ▪ Outils et plates-formes de développement, présentation de Jena ▪ Protégé ▪ Installer et configurer l'environnement de développement ▪ Créer et lire un modèle RDF ▪ Gérer des modèles RDF persistants 	2h
Triple store <ul style="list-style-type: none"> ▪ Open RDF ▪ Sesame triplestore : installation- contraintes de production ▪ JENA Triplestore : installation – scripts ligne de commande - configuration ▪ Virtuoso Triplestore : Installation – fonctionnalités spécifiques 	2h
Cycle de vie des données / versionning <ul style="list-style-type: none"> ▪ Évolution temporelle des données 	30min
Développement et utilisation des services sémantiques <ul style="list-style-type: none"> ▪ Implémenter des services sémantiques <ul style="list-style-type: none"> - Exposer des Services Web XML en tant que RDF - WSMO : Web Service Modeling Ontology - Annotations sémantiques pour WSDL - SAWSDL : exemples et outils <i>Travaux pratiques : mettre en œuvre un service sémantique</i>	3h
Intégration - exploitation d'une plate-forme exposant des services sémantiques - Contraintes Risques <ul style="list-style-type: none"> ▪ Définir et peupler le modèle avec des données sémantiques ▪ Comment exposer un service sémantique :? Problématique à traiter : intégration de contenus hétérogènes, indexation, mise à jour, mise à l'échelle ▪ Risques : Les analyses à mener ▪ Évaluer les contraintes d'indexation ▪ Comment faire évoluer une plate forme sémantique (évolution du modèle, de l'ontologie,...) 	2h
Utilisation de ressources du web des données <ul style="list-style-type: none"> ▪ Les différentes approches et limites dans la réutilisation de ressources du web des données: <ul style="list-style-type: none"> - utilisation du endpoint sparql - utilisation des fichiers mis à disposition - utilisation des ressources d'alignement - la tracabilité des données réutilisées <i>Travaux pratiques : réutilisation de données de fr.dbpedia, réutilisation de données de data.bnf.fr</i>	3h
Les moteurs de recherche <ul style="list-style-type: none"> ▪ Interface utilisateur ▪ Les différents modules utilisés dans les moteurs actuels ▪ Les avancées en matière de recherche en langage naturel 	1h
Veille <ul style="list-style-type: none"> ▪ Sites, listes de diffusion, blogs à connaître 	1h
Total	5 jours

Web 3.0 : Professionnels de l'information-documentation

CONTEXTE

Le Web 3.0 entraîne une évolution des pratiques et des outils qui nécessite de sensibiliser les décideurs et d'accroître la compétence des différents acteurs par des actions de formation. Pour développer une culture commune du changement et de l'innovation autour du Web sémantique, il apparaît nécessaire de mettre en place un plan de formation interdisciplinaire et inter-institutionnel, qui intègre l'ensemble des acteurs de cette évolution et permette la montée en compétences des différents acteurs des projets du Web 3.0.

Parmi les acteurs, les spécialistes de l'information et de la documentation sont impliqués dans le choix des contenus, la définition de nouveaux usages, la modélisation conceptuelle, l'identification, la consolidation et la gestion des référentiels. L'acquisition de la maîtrise des techniques du web sémantique s'inscrit notamment dans la continuité du travail et des savoir-faire des documentalistes. Le Web 3.0 construit ainsi des modèles pour structurer et lier des informations qui ont pour but d'être partagées et s'appuie notamment sur la structuration et la gestion de vocabulaires et de schémas de données et ouvre de nouvelles possibilités de valorisation des contenus et le développement de services. En cela il mobilise des compétences déjà intégrées par les professionnels de l'information.

Cependant si les documentalistes ont développé des méthodologies, des terminologies et des outils pour inventorier et cataloguer les fonds documentaires dont ils ont la gestion, l'objectif est, à présent, d'engager ce travail à l'échelle du web, et non plus dans le cadre d'un fonds documentaire fermé, et d'ouvrir ainsi leur cadre d'intervention et le champ de leur réflexion. De même, les professionnels de l'information-documentation doivent intégrer dans leur offre de ressources et de services d'information les nouvelles pratiques induites par la facilité d'accès, de partage et de diffusion, et les apports du web sémantique. Ils ont également un rôle important, au cœur de leur offre de services à destination de leurs publics, de conseil et d'expertise, mais également dans le transfert de connaissances et la diffusion des bonnes pratiques liées au Web 3.0.

Un plan de formation adapté aux besoins et prenant en compte les spécificités « métier » doit amener les documentalistes vers un changement de modèle et l'acquisition de nouveaux comportements.

PUBLIC CIBLE

L'ensemble des professionnels de l'information-documentation (documentalistes, bibliothécaires, administrateurs de données, gestionnaires de thésaurus, indexeurs, veilleurs,...)

PREREQUIS EVENTUELS

- Connaître les notions de base du web sémantique et du web des données ainsi que les principaux éléments de vocabulaire

OBJECTIFS GENERAUX DE LA FORMATION

- Comprendre les enjeux et les apports des technologies du web 3.0 en termes d'évolution « métier »
- *Comprendre les enjeux stratégiques pour le MCC*
- Permettre d'initier leur mise en œuvre dans un projet. Cet objectif peut se décliner en trois points :
 - faire comprendre le web sémantique : langages, techniques, apports et valeur ajoutée, notamment dans le cadre de l'évolution des bases de données et des fonctions de recherche (veille), de capitalisation et de valorisation de ressources documentaires ;
 - donner les clés de compréhension permettant de mettre à disposition et faire évoluer des bases de données en utilisant les technologies du web 3.0 ;
 - permettre d'appréhender ce que sont les ontologies, leur intérêt pour les applications actuelles ou en projet, les moyens de les mettre en œuvre.

OBJECTIFS PEDAGOGIQUES

À l'issue de ce stage, les participants doivent :

- connaître les principaux concepts, standards du Web 3.0 (URI, RDF, RDFS, OWL, SPARQL) et les modèles couramment utilisés (FOAF, SKOS, FRBR) ainsi que les outils permettant de les mettre en œuvre
- être capables de :
 - définir les principaux éléments d'une stratégie adaptée à leurs objectifs
 - définir une démarche de constitution de ressources dans le contexte web 3.0
 - reconnaître les opportunités du web 3.0 pour faire évoluer leurs applications vers un contexte web sémantique
 - accompagner, présenter et être des relais de la connaissance liée au web 3.0 auprès des autres agents
 - initier des méthodologies pour construire des ontologies

ELEMENTS DU PROGRAMME

Thèmes/contenus	Durée estimée
Cadre général <ul style="list-style-type: none"> ▪ Enjeux, principes (interopérabilité, etc.), définition des notions clefs (LOD, Open data 5 étoiles, etc.) 	45 mn
Exemples d'applications et d'interfaces utilisateurs <ul style="list-style-type: none"> ▪ Exemples d'applications et d'outils sur le Web (Knowledge Graph de Google, Relfinder..) ▪ <i>Panorama des réalisations du MCC et de ses EP (Data.bnf, JocondeLab, Ginfo...)</i> ▪ La transformation des modalités d'accès; penser l'accès pour les utilisateurs et l'accès pour les machines ▪ Panorama des offres progiciels 3.0 	2h
Identifiants / URI <ul style="list-style-type: none"> ▪ Négociation de contenu : Différence entre URL et URI ▪ Une ressource plusieurs représentations ▪ Identifiants pérennes ▪ Règles de nomage des URI pour les différents types de ressources ▪ Redirection des URIs 	30 mn
Triplets et RDF (Resource Description Framework) <ul style="list-style-type: none"> ▪ Notion de triplet, syntaxe et sémantique du modèle RDF ▪ Présentation détaillée et apports du modèle de données RDF ▪ Comparaison base de données relationnelles et RDF 	2 h
RDF Schema <ul style="list-style-type: none"> ▪ Présentation des schémas RDFS et éléments du langage RDFS ▪ Pourquoi associer un schéma RDFS à une description RDF? ▪ Créer un schéma RDFS 	2h

<p>Ontologies</p> <ul style="list-style-type: none"> ▪ Qu'est-ce qu'une ontologie : représentation des connaissances, syntaxe, sémantique ▪ Web Ontology Language (OWL) ▪ Pourquoi développer une ontologie dans une démarche Web sémantique ▪ Les ontologies et vocabulaires de référence existantes : les trouver, les faire évoluer, l'annuaire LOV ▪ Notions d'ontologies modulaires ▪ Principaux outils techniques de gestion des ontologies ▪ Illustration d'une ontologie au travers d'un modèle <ul style="list-style-type: none"> - Caractérisation des différences entre thésaurus et ontologie - Comment construire une ontologie : <ul style="list-style-type: none"> — assemblage d'ontologies — profil d'application — méthodologie de construction à partir d'un corpus ▪ Notion sur l'inférence 	6h
<p>Langages d'interrogation</p> <ul style="list-style-type: none"> ▪ Présentation du langage d'interrogation SPARQL; fonctionnement, syntaxe et exécution de requêtes, recherche d'informations 	2h
<p>SKOS: Terminologie et alignement</p> <ul style="list-style-type: none"> ▪ SKOS : présentation détaillée de SKOS, modèle de données pour vocabulaire contrôlé (thésaurus, autorités, référentiels...) et lien avec la norme ISO 25964-1 ▪ Mise en œuvre, migration au format SKOS, export-import, logiciels utilisant le format SKOS ▪ Terminologie locale vs terminologie de référence ▪ Alignement de vocabulaires (SKOS) ▪ Les outils d'aide à l'alignement (ex. ONAGUI, EDOAL...) ▪ Lemon : la modélisation de données lexicales 	4h
<p>FRBR: Notice bibliographique</p> <ul style="list-style-type: none"> ▪ Présentation du modèle de catalogue bibliographique FRBR / FRBR-OO 	2h
<p>Cidoc CRM</p>	2h
<p>FOAF</p>	30 min
<p>Schema.org : Référencement sémantique des pages HTML</p> <ul style="list-style-type: none"> ▪ Des pages web annotées sémantiquement : Schema.org, microdata et RDFa ▪ L'impact de l'annotation sémantique des pages web ▪ Exemples d'utilisation 	2h
<p>Cycle de vie et méthodologie de projets</p> <ul style="list-style-type: none"> ▪ Intégrer les besoins des différents types d'utilisateurs : site web, données réutilisables, web services ▪ Le dialogue avec les chefs de projet MOA/MOE et la perspective utilisateur ▪ Spécificités méthodologiques d'un projet web sémantique 	3h
<p>Veille</p> <ul style="list-style-type: none"> ▪ Sites, listes de diffusion, blogs à connaître 	1h
Total	5 Jours