

Séminaire Axe 4 Gériico

Humanités Numériques : Interopérabilité et Web sémantique, métadonnées schémas et annotation sémantique

Maison de la Recherche, salle des colloques 14h00-17h30h

Programme

14h00 Introduction

14h10-14h40 "Les métadonnées à l'appui de la pérennisation de l'information numérique : la gestion de données dans SPAR, le magasin numérique de la BnF"

Bertrand Caron, Bibliothèque nationale de France, Département des métadonnées

Résumé : "Depuis 2010, la BnF préserve ses collections numériques acquises ou reçues par dépôt légal dans un magasin numérique, SPAR, dont la base logicielle a été développée pour ses besoins propres. Avec une volumétrie prévisionnelle très conséquente (actuellement 4 pétaoctets de données, en augmentation rapide), l'institution a précocement pris conscience du besoin d'appuyer les stratégies de pérennisation de l'information numérique sur une connaissance approfondie de ses collections. Pionnière parmi les institutions culturelles dans les technologies du web sémantique, la BnF a donc utilisé un entrepôt de triples comme base de données pour son magasin numérique. La puissance et la flexibilité de ces technologies, tout en présentant des difficultés liées à leurs performances et à leur appropriation, permettent de lier les données concernant les objets numériques, leur provenance, leurs dépendances, mais aussi les politiques de préservation qui leur sont affectées. Récemment, les efforts pour proposer aux agents des méthodes d'accès plus intuitives aux données du magasin viennent s'ajouter aux outils du gestionnaire de collection numérique et démontrent une fois de plus les bénéfices de l'ouverture des données pour l'institution elle-même.

14h40-15h10 "La vie des données : de la constitution du corpus à l'édition électronique des écrits d'André-Marie Ampère"

Philippe Pons, Collège de France, Centre de recherche sur les civilisations de l'Asie orientale

Résumé : Comment éditer en ligne un corpus hétérogène constitué de publications, de correspondance et de manuscrits ? C'est ce que nous proposons de présenter en insistant particulièrement sur le travail de transcription, d'encodage en XML-TEI et de mise en ligne qui ont permis d'aboutir à une nouvelle édition électronique des écrits du savant André-Marie Ampère.

15h10-15h30: Discussion

15h30-15h45: Pause

15h45-16h15 "Retour sur la construction d'un dispositif sociotechnique basé sur un graphe de connaissances à partir de schèmes d'utilisation"

Antoine Henry, Daylight Consulting Company, Docteur en SIC

Résumé : Dans le cadre du séminaire de l'axe 4 : Circulation de l'information et organisation des connaissances et sa thématique « Humanités Numériques : Interopérabilité et Web sémantique, métadonnées schémas et annotation sémantique », notre contribution se présentera sous la forme d'une restitution d'un projet réalisé dans le cadre de nos travaux doctoraux. Ce projet (Tuilage) propose de relier les données pour construire de la valeur pour l'entreprise et ses membres. Si celui-ci ne va pas au bout de la logique technique du Linked Data avec des données décrites en Ressource Description Framework (RDF) (Breslin et al. 2010 ; Hauptmann, Brocco et Wörndl, 2015) et une ontologie, l'approche par le graphe est une première avancée pour tendre vers un traitement sémantique des données. Ce projet autorise une réflexion sur un cas concret dans une organisation où les détenteurs de données sont confrontés aux incohérences, à l'absence de contexte (Powell et Hopkins, 2015), aux différences de structure dans les bases de données, situation qui est assez éloignée du web des données tel qu'il a été imaginé par Berners-lee, Hendler et Lassila (2001). Le dispositif sociotechnique Tuilage s'appuie sur un modèle de connaissance basé sur un graphe de connaissances (knowledge graph) (Arenas et al. 2015 ; Le-Phuoc et al., 2016 ; Rospocher et al., 2016 ; Verborgh et al., 2016) à l'instar de ce que pratiquent Google et Facebook (graph search). Si pour Rospocher et al. (2016) le graphe de connaissance est particulièrement utilisé par les moteurs de recherche pour agréger des données structurées provenant de sources fournissant des données liées comme Dbpedia,

Yago ou Wikidata, dans le cas de l'organisation étudiée, sont utilisées pour l'instant uniquement des ressources internes à l'entreprise.

16h15-16h45 “Cultural interoperability in subject heading schemes: A case of human rights concepts”

Tan Tran, doctorant Geriico

Abstract: Interoperability is a crucial factor to the management of digital repositories, not only on the technical level (architecture, metadata standards, copyright) but also on the social and cultural level (language, collective memory, social attitudes, preferences and practices). In this sense, cultural interoperability reflects the experience of change, which is a process of transformation that involves new knowledge objects, forms of alignment, and behaviours emerging on the Web. This presentation provides an analysis of cultural interoperability applied in human rights subject representation in a wide range of aspects concerning linguistic and cultural diversity, cultural warrant and cultural hospitality, marginalization and exclusion as problems of lacking cultural interoperability.

16h45-17h15 : Discussion

Short Bios

Bertrand Caron est archiviste paléographe. Il a occupé de 2011 à 2014 un poste de chef de projet numérisation patrimoniale à la BIU de Montpellier. Il est depuis 2014 expert de modélisation au Département des métadonnées de la BnF et travaille sur la description des objets numériques en vue de leur préservation à long terme. Membre des comités éditoriaux internationaux METS Editorial Board et PREMIS Editorial Committee, il contribue à l'évolution des standards de métadonnées pour les adapter aux technologies du web sémantique."

Antoine Henry, is Senior Researcher at the Aurore program Daylight Consulting Company. He holds a PhD degree from Aix Marseille University on the transformation of the French national gas distributor regarding the digital instrumentation for sociotechnical purposes based collective intelligence to lead to the concept of 'Organizational eXperience' (OX). His main areas of interest are information, collective intelligence, and organization transformation and information system. Besides, he is an associated teaching staff member at the

University Paris 4 Dauphine University, where he is charge of courses and seminars on project management and digital strategy.

Philippe Pons est titulaire d'un Master Histoire et Nouvelles Technologies de l'école nationale des chartes. Il a d'abord travaillé des sources médiévales, comme l'édition de la première traduction de la Cité de Dieu de Saint Augustin au XIV^{ème} siècle par Raoul de Presles et a ensuite participé au projet ANR Ampère. Il est aujourd'hui ingénieur d'étude chargé de l'édition de corpus numérique au Centre de recherche sur les civilisations de l'Asie orientale (UMR 8155 CNRS, EPHE-PSL, Collège de France, Université Paris Diderot). Dans ce cadre, il travaille sur plusieurs projets en Humanités Numériques en rapport avec l'Asie orientale, comme le Dictionnaire de l'agriculture traditionnelle multilingue (français-anglais-chinois-japonais) disponible à l'adresse suivante: <http://dicat.huma-num.fr>.

Quoc-Tan Tran holds a Master of Research in Library and Information Science from the University of Lille in 2016. He served as systems designer for Smart Art project whose aim is to build a culture-based information architecture for Aboriginal art-making communities in Australia from a design anthropology perspective. He also developed a thesaurus on Jewish heritage to support automatic indexing of photographic artworks in Galicia Jewish Museum (Poland). His areas of investigation include ethics and interoperability in e-infrastructures, and Internet ecologies of open knowledge.