

Gériico

 **Université
de Lille**

LE WEB SÉMANTIQUE ET LE PATRIMOINE CULTUREL

**DE LA CONVERGENCE DES DONNÉES AU
CROISEMENT DES CONNAISSANCES**

**RÉSUMÉS
(VERSION FRANÇAISE)**

Organisé par

**Kaouther AZOUZ, Dr. GERiico
Laure Bolka-Tabary, MCF - GERiico**

RÉSUMÉS

Atelier. 1 Multilinguisme, dispositifs d'information et web de données (Modérateur - Pr. Stéphane CHAUDIRON)

- La présentation met en avant le recours aux techniques du web de données dans le développement d'une ontologie des noms arabes classiques pour le patrimoine culturel. Le processus sera décrit les différentes phases séquentielles de : spécification, modélisation, génération, mise en relation, publication et exploitation des données -Sherine Eid - Bibliothèque Alexandrina
- Chaque jour, l'humanité crée 2,5 quintillions ($2,5 * 10^{30}$) octets de données. Dans une étude d'IBM en 2017, il ressort qu'au cours des deux dernières années seulement, 90% des données mondiales ont été créées. Avec l'arrivée de nouveaux utilisateurs, services, appareils, capteurs et technologies, le taux de croissance des données est susceptible de s'accroître. L'évolution du Web traditionnel vers le Web sémantique permet à la machine d'être un citoyen de premier ordre sur le Web et augmente la découverte et l'accessibilité des données non structurées sur le Web, y compris les informations sur le patrimoine culturel. Cette évolution permet à la technologie des données liées d'être utilisée comme base de connaissances pour les données non structurées, notamment les textes, aujourd'hui disponible sur le Web. Plus encore, cette explosion de données doit être accessible aux humains qui ont besoin d'informations et de réponses à leurs questions. Le meilleur scénario est celui des systèmes questions-réponses (QAS); ce type de système permet aux utilisateurs non professionnels (utilisateurs ordinaires) de poser une question en langage naturel et de renvoyer la bonne réponse à leur question à partir des données liées ". - Abdelghani Bouziane - University of SBA

Atelier. 2 Retour d'expérience sur l'impact du web sémantique en bibliothèques (Modérateur - Eric KERGOSIEN)

- L'auteur s'intéresse à la manière de mise en œuvre des données ouvertes liées (Linked Open Data) dans les institutions culturelles polonaises. La présentation aborde les résultats d'une enquête nationale (janvier 2021) auprès des bibliothécaires, muséologues et archivistes polonais concernant leur connaissance des LOD et de leur existence. L'impact sur l'interopérabilité des systèmes d'organisation de la connaissance créés par les bibliothèques et autres institutions est également présenté en se basant sur les plus grands projets des bibliothèques polonaises qui utilisent LOD - Dorota Siwecka - Uniwersitet Worclawski
- Cette communication prendra la forme d'un retour d'expérience sur le projet data.bnf.fr, lancé par la Bibliothèque nationale de France en 2011. Ce service apparaissait comme une réponse concrète à un double constat. Premièrement, les données de la BnF étaient alors difficilement accessibles sur le web puisque peu référencées par les moteurs de recherche. Deuxièmement, la publication du modèle conceptuel FRBR (puis IFLA-LRM) proposait une nouvelle structuration des données plus appropriée au contexte de développement du web et incitait à s'emparer des technologies du web sémantique. En pratique, data.bnf.fr permet d'agrégier sur une même page des données provenant de différents réservoirs de la BnF, tout en affichant des liens vers d'autres entités et vers des données produites par d'autres opérateurs culturels et patrimoniaux. En outre, les données sont structurées dans des pages HTML de telle sorte qu'elles soient identifiées et trouvables efficacement par l'intermédiaire des moteurs de recherche. Elles sont enfin diffusées en RDF, en utilisant des ontologies qui font universellement référence. 10 ans après, le projet a atteint une volumétrie importante (près de 4 millions de pages) ; il reste néanmoins en évolution constante. Nous présenterons ainsi quelques perspectives de travail pour les prochaines années, en mettant notamment l'accent sur le travail de médiation qu'implique l'objectif initial de simplification de l'accès aux données.- Tiphaine Cécile Foucheret Xavier Levoin - Bibliothèque nationale de France

Atelier. 3 Paradigmes du web sémantique sur les communautés expertes et la connaissance (Modératrice - Marianne CAILLOUX)

- Europeana rassemble des métadonnées pour donner accès à des dizaines de millions d'objets numérisés qui proviennent des collections des bibliothèques, archives et musées répartis dans toute l'Europe. Pour résoudre les problèmes de qualité de données qui sont apparus dès les débuts de l'initiative, et essayer de fournir un service qui fonctionne mieux à travers les frontières institutionnelles et linguistiques, Europeana et ses partenaires se sont tournés vers les technologies du web de données. Dans cette présentation, nous allons détailler comment nous avons employé différents aspects de ces technologies pour améliorer la représentation des métadonnées, leur interopérabilité, et leur richesse, en particulier dans un contexte multilingue. Nous illustrerons l'impact de cet emploi, par exemple pour les utilisateurs du portail europeana.eu. Au-delà des éléments techniques, nous aborderons aussi le rôle crucial de la communauté dans le succès des efforts entrepris, ainsi que l'importance de suivre (et promouvoir) les bonnes pratiques qui gravitent autour des solutions techniques.- Antoine Isaac - Europeana

RÉSUMÉS

- Le projet Médicis est un projet de recherche universitaire pluridisciplinaire du territoire culturel Centre Val de Loire. Médicis propose de reconstituer le cadre architectural et ses décors mobiliers du château d'Amboise à l'époque de la Renaissance afin que les habitants du Centre Val de Loire puissent se réapproprier le mobilier des châteaux dans un contexte d'identité ligérienne. Le développement de ce projet fait appel à plusieurs acteurs de la culture, historiens et spécialistes du mobiliers et plusieurs laboratoires universitaires. Cette recherche universitaire est menée sur trois années au service de la fondation Saint-Louis qui gère le château d'Amboise et le mobilier des châteaux ligériens royaux pour permettre d'améliorer la reconnaissance du mobilier de l'époque et de transposer cette connaissance à d'autres demeures princières du Val-de-Loire. Les données produites dans le cadre du projet Médicis devront alimenter plusieurs plateformes de données patrimoniales et permettre de voir émerger un dispositif numérique au sein du château qui vise à l'immersion du grand public dans les décors reconstitués. En effet, si les châteaux sont des lieux incontournables du patrimoine ligérien, ils sont aujourd'hui, tout comme les musées, perçus comme des établissements d'expérience culturelle et de communication envers le public qu'ils reçoivent. Cette communication aura pour objectif dans un premier temps d'exposer le projet Médicis et ses objectifs de médiation patrimoniale. Dans un second temps, je reviendrais sur la méthodologie de la collecte des données du projet en rappelant l'historique des processus documentaire du musée gestionnaire qui a permis de cataloguer toutes les données aujourd'hui en ligne. Enfin, à l'aide des dernières recherches en SIC sur la médiation en musée, il s'agira de pointer les enjeux du dispositif numérique imaginé mais également d'en montrer les limites. - Maryse Rizza - Université de Tours

Atelier. 4 De la convergence des données à la réorganisation des connaissances (Modératrice - Joana CASENAVE)

- Depuis 2013, le groupe d'experts EGAD du Conseil international des Archives travaille à l'élaboration d'un nouveau standard pour la description des archives, Records in Contexts (RiC). Ce standard, dont la version 1.0 sera publiée fin 2021, se composera alors de trois parties : un modèle conceptuel entités-relations, RiC-CM ; une ontologie OWL, RiC-O et un manuel d'application, RiC-AG. La version 0.1 de RiC-O a été publiée en décembre 2019, sa version 0.2 fin janvier 2021 (<https://www.ica.org/standards/RiC/ontology>). De même que LRM et RDA pour les bibliothèques, RiC a pour ambition d'accompagner les archivistes dans la transition des métadonnées archivistiques vers des graphes d'entités, et RiC-O de leur permettre de produire des jeux de données RDF de qualité, pour les publier dans le web de données et les lier à d'autres jeux de données. Les Archives nationales de France ont expérimenté l'utilisation de RiC-O dès 2015, et sont récemment passées à l'échelle, en produisant en masse des jeux de données RDF/RiC-O et en commençant à travailler à l'enrichissement de leurs référentiels et à leur alignement avec d'autres jeux de données. Ces travaux ont donc produit de premiers résultats concrets (un prototype en ligne réalisé en collaboration avec la Bibliothèque nationale de France et le ministère de la Culture, PIAAF ; un logiciel libre, RiC-O Converter ; des jeux de données massifs dont certains sont déjà utilisés). En 2021 ils vont se poursuivre. Un bilan de ces premiers travaux sera rapidement présenté ainsi que les perspectives pour 2021, notamment celles qui impliquent des collaborations extérieures. Si ce programme vise à inscrire les Archives nationales de France dans l'écosystème du web de données pour le bénéfice des utilisateurs et des machines, il a aussi permis d'initier la transition mentionnée ci-dessus, dont les impacts seront à terme profonds sur les pratiques descriptives et sur le système d'information de l'institution. La présentation évoquera certaines dimensions de cette stratégie, en particulier celle de la montée en qualité des métadonnées. - Florence Clavaud - Archives nationales de France
- Présentation d'une méthodologie composite permettant de créer des ontologies de domaines dans le cadre de la préservation et la valorisation du patrimoine culturel et industriel. Fondée sur la reprise d'outils documentaires divers, l'analyse terminologique, des entretiens ethnographiques, la fouille de textes et l'analyse cartographique du web, cette approche est expérimentée dans trois univers distincts : la musique, le textile et les mines de charbon.- Bernard Jacquemin - GERiCO - Université de Lille
Mots-clés : ontologie, musique, textile, mines de charbon.

GERiico

 **Université
de Lille**

THE SEMANTIC WEB AND CULTURAL HERITAGE

**FROM DATA CONVERGENCE TO
KNOWLEDGE CROSSING**

**ABSTRACT
(ENGLISH VERSION)**

Organised by

**Kaouther AZOUZ, PhD. GERiCO
Laure Bolka-Tabary, Lecturer - GERiCO**

ABSTRACT

Workshop. 1 Multilinguism, information devices and Linked data (Moderation - Pr. Stéphane CHAUDIRON)

- The presentation identifies ontology engineering phases to develop a proposed linked data ontology for Arabic classical names in the domain of cultural heritage. The process will be outlined through sequential phases including specification, modeling, generation, linking, publication, and exploitation. Sherine Eid - Bibliothèque Alexandrina
- "Every day, humanity creates 2.5 quintillions ($2.5 * 10^{30}$) bytes of data. In 2017 IBM study outline that in just the last two years, 90% of the world's data was created. With the advent of new users, services, devices, sensors, and technologies, the rate of data growth is likely to accelerate. The evolution of the traditional Web towards the semantic Web allows the machine to be a first-order citizen on the Web and increases the discoverability of and accessibility to the unstructured data on the Web including information on cultural heritage. This evolution enables the Linked Data technology to be used as background knowledge bases for unstructured data, notably the texts, available nowadays on the Web. Even more, this explosion of data must be accessible for humans who need information and a response to their questions. The best scenario is Question Answer Systems (QAS); this kind of system allows lay-users (ordinary users) to ask a question in natural language and return the correct answer to their question over linked data". -- Abdelghani Bouziane - University of SBA

Workshop. 2 Outcomes and feedback of the semantic web's implementation in libraries (Moderation- Eric KERGOSIEN)

- The author presents how does the Linked Open Data implementation look like in Polish cultural institutions landscape. The presentation shows the results of national survey (January 2021) among Polish librarians, museologists and archivist concerning their awareness of the existence of LOD. The impact on the interoperability of knowledge organisation systems created by libraries and other institutions is also presented basing on the largest projects of Polish libraries which uses LOD. - Dorota Siwecka - Uniwersitet Worclawski
- This communication will provide feedback on the data.bnf.fr project, launched by the National Library of France in 2011. This new service was a concrete answer to a twofold issue. Firstly, BnF data were not easily accessible on the web, since they were not very well referenced by search engines. Secondly, the publication of the FRBR conceptual model (then IFLA-LRM) offered a new data organization more suited to the context of the web's development and encouraged the use of semantic web technologies. In practical terms, data.bnf.fr aggregates on the same page data from different BnF databases, while displaying links to other entities and to data produced by other cultural and heritage providers. In addition, data is structured in HTML pages so that it can be identified and found efficiently through search engines. Finally, data is published in RDF, using ontologies of universal reference. 10 years later, the project has reached a significant volume (nearly 4 million pages); it is nevertheless in constant progress. We will thus present some work perspectives for the coming years, with particular emphasis on the mediation work involved in the initial objective of simplifying data access.- Tiphaine Cécile Foucher et Xavier Levoine - National Library of France

Workshop. 3 Semantic web paradigm for Knowledge and expert communities (Moderation - Marianne CAILLOUX)

- Europeana gathers metadata to provide access to dozens of millions of digitized objects from libraries, archives, and museums across Europe. To solve data quality issues that have occurred since the beginning and provide a service that works well across institutional and language barriers, Europeana and its partners have sought to re-use linked data technology. In this talk, we will detail how we have employed various components of this technology to enhance metadata modeling, interoperability, and richness, especially considering our massively multilingual context. We will illustrate the impact of these efforts, for example in terms of the user experience on Europeana's portal. Beyond technical aspects, we will also touch on the crucial role of community work for the success of such endeavors and emphasize the importance of following and promote best practices on the use of technology. - Antoine Isaac - Europeana

ABSTRACT

- Heritage documentation and open data: How to reconstruct the hypothesis of the construction of knowledge around a work in a digital device: the case of the Médicis project. Maryse RIZZA PRIM Laboratory University of Tours Summary The Médicis project is a multidisciplinary university research project of the Centre Val de Loire cultural territory. Médicis proposes to reconstruct the architectural framework and its furniture decorations of the castle of Amboise at the time of the Renaissance so that the inhabitants of the Centre Val de Loire can reclaim the furniture of the castles in a context of Ligerian identity. The development of this project involves several cultural actors, historians and specialists in furniture and several university laboratories. This academic research is conducted over three years in the service of the Saint-Louis who manages the « Château d'Amboise » and proposes to collect a set of data on the furniture of the royal Ligerian castles to improve the recognition of the furniture of the time and to transpose this knowledge to other princely residences of the Val-de-loire. The data produced as part of the Médicis project will feed several heritage data platforms and allow for the emergence of a digital device within the castle that aims to immerse the public in the reconstituted decors. Indeed, although castles are essential places of Ligerian heritage, they are today, just like museums, perceived as institutions of cultural experience and communication to the public they receive. The initial objective of this communication will be to present the Médicis project and its heritage mediation objectives. In a second step, I would return to the methodology of data collection of the project by recalling the history of the documentary processes of the museum manager which made it possible to catalogue all the data today online. Finally, using the latest research in CIS on museum mediation, it will be a matter of pointing out the stakes of the imagined digital device but also of showing its limits. - Maryse Rizza - University of Tours

Workshop. 4 From data convergence to Knowledge reorganisation (Moderation - Joana CASENAVE)

- Since 2013, the EGAD expert group of the International Council on Archives has been working on the development of a new standard for describing archives, Records in Contexts (RiC). This standard, version 1.0 of which will be published at the end of 2021, will then consist of three parts: a conceptual entity-relationship model, RiC-CM; an OWL ontology, RiC-O, and application guidelines, RiC-AG. RiC-O version 0.1 was released in December 2019, version 0.2 at the end of January 2021 (<https://www.ica.org/standards/RiC/ontology>). Like LRM and RDA for libraries, RiC aims to support archivists in the transition of archival metadata to entity graphs, and RiC-O to enable them to produce good quality RDF datasets, for publishing them into the web of data and linking them to other datasets. The National Archives of France experimented with the use of RiC-O as early as 2015, and recently scaled up, producing big RDF/RiC-O datasets and starting to enrich them and align them with other datasets. This work therefore produced first concrete results (an online prototype produced in collaboration with the National Library of France and the Ministry of Culture, PIAAF; a free software, RiC-O Converter ; and datasets, some of which are already used). In 2021 they will continue. An assessment of this initial work will be quickly presented as well as the prospects for 2021, in particular those involving external collaborations. While this program aims to include the National Archives of France in the ecosystem of the web of data for the benefit of users and machines, it has also made it possible to initiate the transition mentioned above, whose impacts will ultimately be profound on the institution's description practice and methods and on its information system. The presentation will discuss some facets of this strategy, in particular that of improving the quality of metadata. - Florence Clavaud - National Archives of France
- Presentation of a composite methodology for creating domain ontologies in the context of the preservation and enhancement of cultural and industrial heritage. Based on the use of various documentary tools, terminological analysis, ethnographic interviews, text mining and cartographic analysis of the web, this approach is experimented in three distinct universes: music, textiles and coal mines - Bernard Jacquemin - GERiiCO - Université de Lille