

Le web de données (linked data) pour le partage des connaissances des bibliothèques, archives et musées (LAM)

Synthèse opérative

Les bibliothèques, les archives et les musées (LAM) possèdent une grande quantité de données et de ressources qui, jusqu'à présent, sont souvent dissimulées dans les catalogues et les archives. Optimiser le potentiel de valorisation et de divulgation des informations à un large public représente une opportunité d'enrichissement du World Wide Web. La promotion de l'ouverture vers la connaissance permet de créer une série d'avantages pour tous les éléments de la chaîne de l'information. Pour les bibliothèques, archives et musées cela représente une précieuse possibilité de créer un lien entre les collections et de fournir aux usagers des instruments plus complets offrant une grande quantité de données. De plus, les éditeurs sont également en mesure de proposer de nouvelles formes de coopération. Le patrimoine culturel mondial peut ainsi être préservé au bénéfice des générations futures et la recherche peut être maintenue dans la langue originale dans un esprit de conservation de la variété et de la vitalité culturelles. Les bénéfices potentiels apportés dans le secteur des sciences humaines, secteur qui risque plus que jamais la marginalisation, sont significatifs.

Le web de données est un instrument qui rend tout cela possible.

Cette nouvelle approche du traitement des données et de l'information pose les bases pour de nouvelles améliorations dans la manière dont les institutions interagissent entre elles et pour la réutilisation des données dans différents contextes, permettant ainsi une identification et une visibilité des entités Web plus efficaces. Dans le même temps, la nouvelle approche soutient l'utilisation des collections du patrimoine culturel et encourage l'utilisation et le partage des données plus largement. En appliquant le paradigme du web de données, les bibliothèques, les archives, les musées et les professionnels de l'information disposeront d'une suite d'outils plus complète. Ces outils évolueront au fur et à mesure que les institutions s'engageront dans un espace de travail commun et auront la liberté de suivre le rythme de la technologie web à mesure qu'elle subira sa propre évolution.

À la fois flexible et durable, cette approche coopérative permet de créer un système bien structuré d'organisation des données, en tenant compte de diverses traditions culturelles et en exploitant les opportunités technologiques. Les chiffres clés du secteur peuvent contribuer aux côtés de petites organisations qui sont prêtes à partager leur propre expérience et leurs atouts d'exception.

Pour se conformer à la nouvelle philosophie ouverte du partage et de la réutilisation des données, le concept des notices d'autorité traditionnelles doit également évoluer. La discussion précédente sur la question de savoir si le contrôle de l'autorité devrait être centralisé ou local, se transformera au fur et à mesure que l'on passera d'une conception rigide du contrôle de l'autorité à une notion plus flexible d'identification des entités et de relations entre entités. Des organisations telles que les bibliothèques, les archives, les musées, les éditeurs et les fournisseurs s'engagent dans la création de nouvelles données et la découverte de nouvelles ressources, dépassant les limites de leurs domaines spécifiques pour créer des opportunités d'enrichissement de données auparavant impensables.

De même, le nouveau modèle de données BIBFRAME fait actuellement l'objet de nombreuses discussions et développements au sein de la communauté des bibliothèques et au-delà. L'objectif est de développer les possibilités offertes par le web de données aux bibliothèques, archives et musées, en offrant une interopérabilité, une visibilité et une localisation accrues pour tous les types de ressources tout en tenant compte de l'évolution des lignes directrices et des modèles de données comme le modèle IFLA Library Reference Model (LRM) et les lignes directrices Resource Description and Access (RDA).

L'objectif des composants et des outils développés dans le cadre d'initiatives Share basées sur des linked data, incluant Share-VDE (environnement de découverte virtuelle), Share-ART (y compris les bibliothèques d'histoire de l'art du Max Planck Institut) et Share-MUSIC (une expérimentation dans le domaine de la musique), est de créer un environnement utile pour la gestion des connaissances issues de la recherche. Les interfaces de découverte avancées améliorent l'expérience de l'utilisateur et offrent des résultats de recherche plus larges aux usagers des bibliothèques, des archives et des musées, tandis que les bibliothécaires, les archivistes et les conservateurs de musée notent plusieurs avancées: développement continu des fonctions de catalogage natives du Web sémantique, l'intégration des processus dans un contexte collaboratif avec des systèmes et instruments locaux. L'objectif est d'aider à révéler la richesse des données dans les collections existantes, qui restent souvent cachées ou inexprimées dans un catalogue traditionnel.

Au sein du programme, les notices bibliographiques des bibliothèques, archives et musées sont transformées en web de données grâce à l'identification des entités, à la réconciliation et à l'enrichissement des données. Les attributs sont utilisés pour identifier de manière unique une personne, une oeuvre ou d'autres entités, avec des variations réconciliées pour former un groupe de données se référant à la même entité. Ceux-ci sont ensuite rapprochés des fichiers d'autorité traditionnels et d'autres sources externes pour enrichir les documents et créer un réseau d'informations et de ressources. Une base de données de relations, ouverte à l'ensemble de la communauté, est compilée, ainsi qu'une base de connaissances commune de clusters accessible en RDF. La base de données utilise les paradigmes du Web sémantique mais permet également aux bibliothèques participantes de continuer à gérer leurs propres données aussi indépendamment que possible.

Parmi les principaux domaines d'activités:

- enrichissement des notices MARC avec des URI;
- conversion de MARC en RDF en utilisant le vocabulaire BIBFRAME et d'autres ontologies supplémentaires, si nécessaire;
- implémentation d'une plate-forme de découverte virtuelle avec une adaptation du modèle de données BIBFRAME développé pour fournir une option de découverte de linked data;
- création d'une base de données relationnelle et d'un cluster de ressources (Sapientia Cluster Knowledge Base) accessible en RDF;
- implémentation d'outils d'interaction directe avec les données, permettant la validation, la mise à jour, le contrôle et la maintenance à long terme des clusters et des URI identifiant les entités;
- procédures automatique de mise à jour de batch;
- diffusion de batch / automatisation des données aux institutions;
- implémentation progressive d'autres cas d'utilisation dans l'ordre de priorité défini par la communauté, parmi lesquels: catalogage dérivé, catalogage d'origine, API pour ILS, rétro-conversion pour les systèmes locaux d'acquisition et d'administration basés sur les formats MARC.

La collaboration des institutions participantes est encouragée afin que les décisions concernant les étapes ultérieures d'un déploiement à grande échelle dans les communautés des bibliothèques, des archives et des musées puissent être basées sur des preuves solides issues d'ensembles de données. Ces ensembles de données ne sont pas simplement convertis en RDF, ils sont également enrichis avec des identifiants et sont interconnectés. Cela est une condition préalable à l'utilisation dans l'environnement du web de données. Ces processus peuvent ensuite contribuer à construire un modèle réaliste d'activités de l'utilisateur en abordant tous les problèmes ainsi que les avantages potentiels d'un déplacement vers l'environnement de linked data. La fonctionnalité domine la conception du programme, fournissant divers environnements et interfaces utilisateur pour la création de données, l'enrichissement et l'approvisionnement au groupe diversifié de bibliothécaires, professionnels, chercheurs, étudiants et étudiants, englobant un large éventail de besoins.

Casalini Libri (www.casalini.it) est une agence bibliographique qui produit des notices d'autorité et bibliographiques et fournit des livres et des revues aux bibliothèques. L'entreprise offre une variété de services de développement des collections. Elle travaille en collaboration avec les éditeurs et les bibliothèques pour fournir des contenus électroniques.

@Cult (www.atcult.it) est une entreprise de développement de logiciels spécialisée dans la conversion des données pour LD. Elle fournit des outils intégrés pour les bibliothèques et des instruments de découverte qui offrent des solutions technologiques efficaces et innovatrices pour l'amélioration de la gestion des informations et le partage des connaissances.