

# Étapes d'implantation d'un projet de Web de données en milieu documentaire



Marielle St-Germain – marielle.st-germain@umontreal.ca

École de bibliothéconomie et des sciences de l'information - Université de Montréal

## Résumé

Cette recherche a comme objectif, entre autres, d'explorer un ensemble de technologies spécifiques au Web sémantique et au Web de données afin de proposer une liste d'étapes pour la mise sur pied de projets les mettant de l'avant au sein d'institutions documentaires.

Ces étapes visent à faciliter l'appropriation de ces technologies par les professionnels de l'information.

Des recommandations en ce qui a trait aux outils et aux ressources nécessaires sont aussi proposées.

## Quelques concepts-clés

### Web sémantique

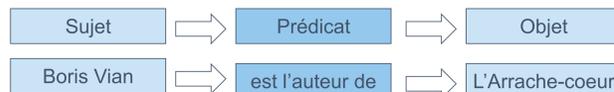
- Extension du Web;
- Vise la publication de données selon certains standards.

### Web de données

- Application du Web sémantique;
- Publication de données en format RDF;
- Création de liens

### RDF

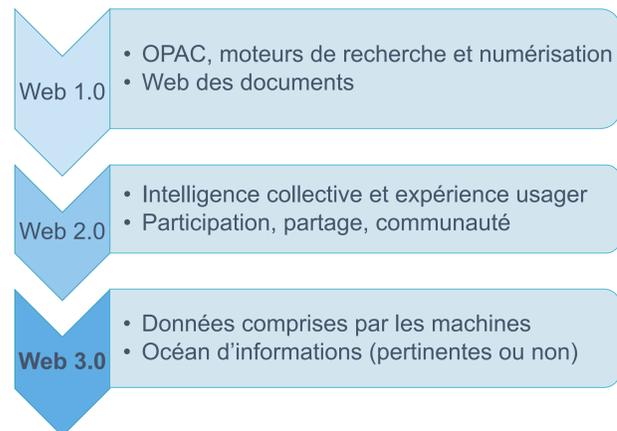
- Langage de base du Web sémantique ayant comme objectif de décrire les ressources et leurs métadonnées;
- Basé sur la formation de triplets.



### URI (identificateur)

- Suite alphanumérique;
- Identifie de manière univoque et pérenne une ressource physique ou abstraite.

## Problématique



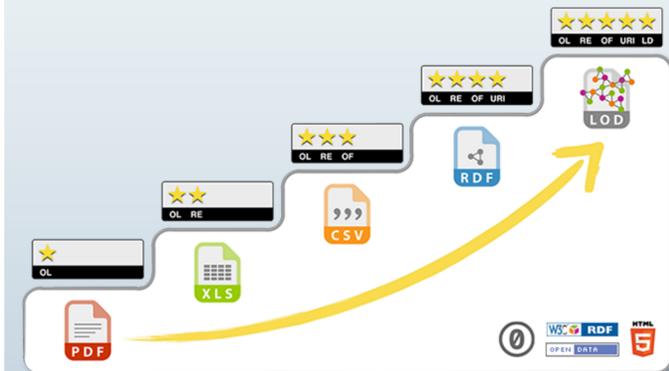
- Gestion de l'information;
- Besoins et comportements informationnels;
- Formats;
- Différences terminologiques entre les bibliothèques et les standards du Web.

## Méthodologie

- Revue de la littérature approfondie;
- Étude des enjeux et de la pertinence du Web sémantique pour les professionnels de l'information;
- Synthèse et regroupement des informations disponibles, décimées dans la littérature;
- Analyse comparative entre la Bibliothèque nationale de France et Bibliothèque et Archives nationales du Québec, en vue d'une éventuelle application au sein de cette dernière institution (entrevues informelles).
- N.B. : À chaque institution ses particularités et l'objectif lié à la présentation des étapes et des pratiques est de proposer une ligne directrice simple pour les institutions.

## Résultats (étapes)

- 1. Comprendre la motivation et prise de conscience;**
- 2. Obtenir l'autorisation des parties prenantes;**
- 3. Établir une licence d'utilisation;**
- 4. Évaluer les compétences;**
  - Systèmes d'information;
  - Métadonnées;
  - Modélisation;
  - Communication.
- 5. Évaluer les jeux de données;**
  - Notices;
  - Éléments de métadonnées ou ontologies;
  - Données administratives.
- 6. Choisir le modèle de publication;**
  - BD relationnelle;
  - API;
  - *Triplestore*;
  - Données en langue naturelle.
- 7. Attribuer les URI;**
  - URI multiples;
  - Identifiants opaques ou signifiants;
  - Pérennité;
  - Négociation de contenu.
- 8. Choisir son modèle de données et faire le mapping;**
  - Vocabulaires, ontologies, schémas.
- 9. Nettoyer les données;**
  - Outils de nettoyage de données.
- 10. Enrichir les données en faisant des liens;**
  - À l'interne et à l'externe (OWL et SKOS).
- 11. Convertir les données en RDF;**
  - Choix de l'outil et des syntaxes de sérialisation.
- 12. Valider les jeux de données;**
- 13. Publier les jeux de données;**
  - Point d'accès SPARQL, *dumps*;
  - API;
  - Données structurées internes.
- 14. Assurer la maintenance du système.**
  - Réévaluation de l'architecture du système;
  - Création de nouveaux liens.



La cote de degré d'ouverture des données de Tim Berners-Lee

## Conclusion

- Défis nombreux, mais pas insurmontables;
- Modèle BIBFRAME 2.0;
- Développement des projets ralenti;
- Nécessité d'établir les pratiques clairement;
- Marches à suivre et outils importants.

## Références

- Bermès, E., Isaac, A. et Poupeau, G. (2013). *Le web sémantique en bibliothèque*. Paris : Electre/Éditions du Cercle de la Librairie.
- Berners-Lee, T., Hendler, J. et Lassila, O. (2001). The semantic web. *Scientific American*, 284(5), 28-37.
- Van Hooland, S. et Verborgh, R. (2014). *Linked data for libraries, archives and museums: How to clean, linked and publish your metadata*. Chicago: Neal-Schuman.

## Remerciements

Cette recherche de maîtrise a été réalisée sous la supervision de Madame Lyne Da Sylva.

Cette recherche a été financée par la Faculté des études supérieures et postdoctorales (FESP) de l'Université de Montréal.