

# Génération de RDF à partir de sources de données aux formats hétérogènes

Maxime Lefrançois, Antoine Zimmermann, Noorani Bakerally

Univ Lyon, MINES Saint-Étienne, CNRS, Laboratoire Hubert Curien UMR 5516,  
F-42023 Saint-Étienne, France  
prenom.nom@emse.fr

**Résumé.** Contrairement à ce que promet le Web des données, les données exposées par la plupart des organisations sont dans des formats non-RDF tels que CSV, JSON, ou XML. De plus sur le Web des objets, les objets contraints préféreront des formats binaires tels que EXI ou CBOR aux formats RDF textuels. Dans ce contexte, RDF peut toutefois servir de lingua franca pour l'interopérabilité sémantique, l'intégration de données aux formats hétérogènes, le raisonnement, et le requêtage. Dans ce but, plusieurs outils et formalismes permettent de transformer des documents non-RDF vers RDF, les plus flexibles étant basés sur des langages de transformation ou de correspondance (GRDDL, XSPARQL, R2RML, RML, CSVW, etc.). Cet article définit un nouveau langage, SPARQL-Generate, qui permet de générer du RDF à partir: (i) d'une base de données RDF, et (ii) d'un nombre quelconque de documents aux formats arbitraires. L'originalité de SPARQL-Generate est qu'il étend SPARQL 1.1, et peut donc (i) être appris facilement par les ingénieurs de la connaissance familiers de SPARQL, (ii) être implémenté au dessus de n'importe quel moteur SPARQL existant, (iii) tirer parti des mécanismes d'extension de SPARQL pour prendre en compte de futurs formats.

## 1 Introduction

Nous cherchons à faciliter l'accès aux formalismes et outils du Web Sémantique pour les entreprises, services Web, et objets contraints. Une étape clé pour utiliser ces formalismes est de générer du RDF à partir de documents ayant des formats variés. En effet, les entreprises et services web stockent et échangent les données dans une multitude de modèles et formats de données : les modèles de données relationnels ainsi que le format XML (pas RDF/XML) sont encore très présent. Les portails open data préfèrent CSV, et les API web : JSON. Quand aux objets contraints sur le Web des objets, ils préfèrent des formats légers, potentiellement binaires, tels que EXI ou CBOR. Dans ce contexte, les formats de données RDF (RDF/XML, Turtle, JSON-LD) ne remplaceront vraisemblablement jamais les formats de données existants. Par contre, le *modèle* de données RDF peut toujours servir de *lingua franca* pour l'interopérabilité sémantique, l'intégration de données aux formats hétérogènes, le raisonnement, et le requêtage.