

**Horizon 2020**  
**INFRADEV-1-2014 - Design studies**

**RICHFIELDS Working Package 11**  
**Deliverable D11.2**

**Semantic data model of the RI Consumer Data  
Platform**

**Date delivered:**  
**M30**

**Authors:**  
Tome Eftimov, Barbara Koroušić Seljak, Gordana  
Ispirova, Peter Korošec, JSI

**Deliverable lead beneficiaries:**  
JSI

<b>Project</b>	
<b>Project acronym:</b>	RICHFIELDS
<b>Project full title:</b>	Research Infrastructure on Consumer Health and Food Intake for E-science with Linked Data Sharing
<b>Grant agreement no.:</b>	654280
<b>Project start date:</b>	01.10.2015
<b>Document:</b>	
<b>Title:</b>	Semantic data model of the RI Consumer Data Platform
<b>Deliverable No.:</b>	D11.2
<b>Authors:</b>	T. Eftimov, B. Koroušić Seljak, G. Ispirova, P. Korošec, IJS
<b>Reviewer:</b>	Karin Zimmermann – Project Coordinator Pieter van 't Veer – Scientific Coordinator Fred van Alphen – Project Advisory Board member
<b>Start date:</b>	1.10.2015
<b>Delivery date:</b>	20.03.2018
<b>Due date of deliverable:</b>	31.03.2018
<b>Dissemination level:</b>	CO
<b>Status:</b>	Final

<b>Change history:</b>		
Version	Notes	Date
001	Draft version	09.02.2017
002	Comments from Krijn Poppe considered	27.02.2018
003	Comments from Fred and Pieter considered	20.03.2018



**Karin Zimmermann**  
**Project Coordinator**



**Prof Pieter van 't Veer**  
**Scientific Coordinator**

## Summary

This deliverable presents concepts and tools for creating a semantic data model for linked (virtually integrated) structured, semi-structured, unstructured, big and open data, based on existing (standard) ontologies and incorporating aspects from the domains of the case studies (WP8-10) as well as data provenance. The summary goes into the need, challenges and methods for making sure we all understand the data in the same way. It is organized as follows: Section 1 presents an explanation of ontology learning process focusing on the basic terms and different processes that can be involved; In section 2, a novel approach for ontology learning is presented, which is the methodology used to learned the RICHFIELDS ontology; Section 3 introduces the text normalization process that is a crucial part of making data machine understandable (i.e. semantic annotation or ontology population), with the use-case with nutrition data; In Section 4, the named-entity recognition process is explained, which is a process of extracting relevant concepts from unstructured data in the domain of interest, together with a use-case of extracting information from evidence-based dietary recommendations; Finally, the conclusions and implications of the provided methodologies to the RICHFIELDS RI are discussed.

Introducing semantics into data, RICHFIELDS has made it possible to embed contextual aspects into the data. To be trustworthy, RICHFIELDS has designed models to describe the provenance of data and explored mechanisms (like digital signatures) to provide the origin and the ownership of linked research data. Considering the user requirements' specification (D11.1), the semantic data model, and the data provenance concept, architecture of the RI Consumer Data Platform was elaborated to enable an integration of all kind of data into linked data. The following services were modelled to support: linkage of data from data pools specified in WP5-7 (consumer generated data) and WP8-10 (business and research generated data), and delivery of linked research data to other information systems, like ELIXIR (<https://www.elixir-europe.org>), CORBEL (<https://www.elixir-europe.org/about/eu-projects/corbel>) and similar.