



THE EIGHTEENTH INTERNATIONAL CONFERENCE ON ADVANCES IN  
SEMANTIC PROCESSING  
SEMAPRO 2024  
VENICE, ITALY

**A SEMANTIC DATA MODEL OF HARMONIZED SURVEY ON HOUSEHOLDS  
LIVING STANDARDS**

Authors

**Marc Mfoutou Moukala<sup>1</sup>, Macaire Ngomo<sup>2</sup>, Régis Freguin Babindamana<sup>1</sup>**

<sup>1</sup> Faculty of Science and Technology, Marien Ngouabi University (Congo)

<sup>2</sup> CM IT CONSEIL, Research and Innovation Department, (France)

Contact email: [moukmarc@yahoo.fr](mailto:moukmarc@yahoo.fr)



## About Marc Mfoutou Moukala

Marc Mfoutou Moukala is currently a doctoral student in Applied Mathematics, specialization in Computer Science at the Faculty of science and technology of Marien Ngouabi University, Congo.

He received a Master's degree in Data Science-Big Data at Institut national polytechnique Félix Houphouët-Boigny (INP-HB) of Yamoussoukro, Côte d'Ivoire in 2020.

He is interested in research and development in Computer science and Artificial intelligence. His research domains include semantic data modeling, knowledge representation, automated reasoning, machine learning and topic modeling.

# 1. Introduction

The Harmonized Survey on Households Living Standards (HSHLS) [1] is the main household statistical survey conducted by French-speaking countries in West and Central Africa since the year 2018, which aims to capture household living conditions information. The results of this survey are used by public authorities and development partners to identify areas where necessary solutions can be provided to households or communities.

However, during data cleaning works, discrepancies are sometimes observed between the methodology and the actual collected data. Moreover, data processing teams are often forced to manually check methodologic documents for analysis purposes, this is a time-consuming task.

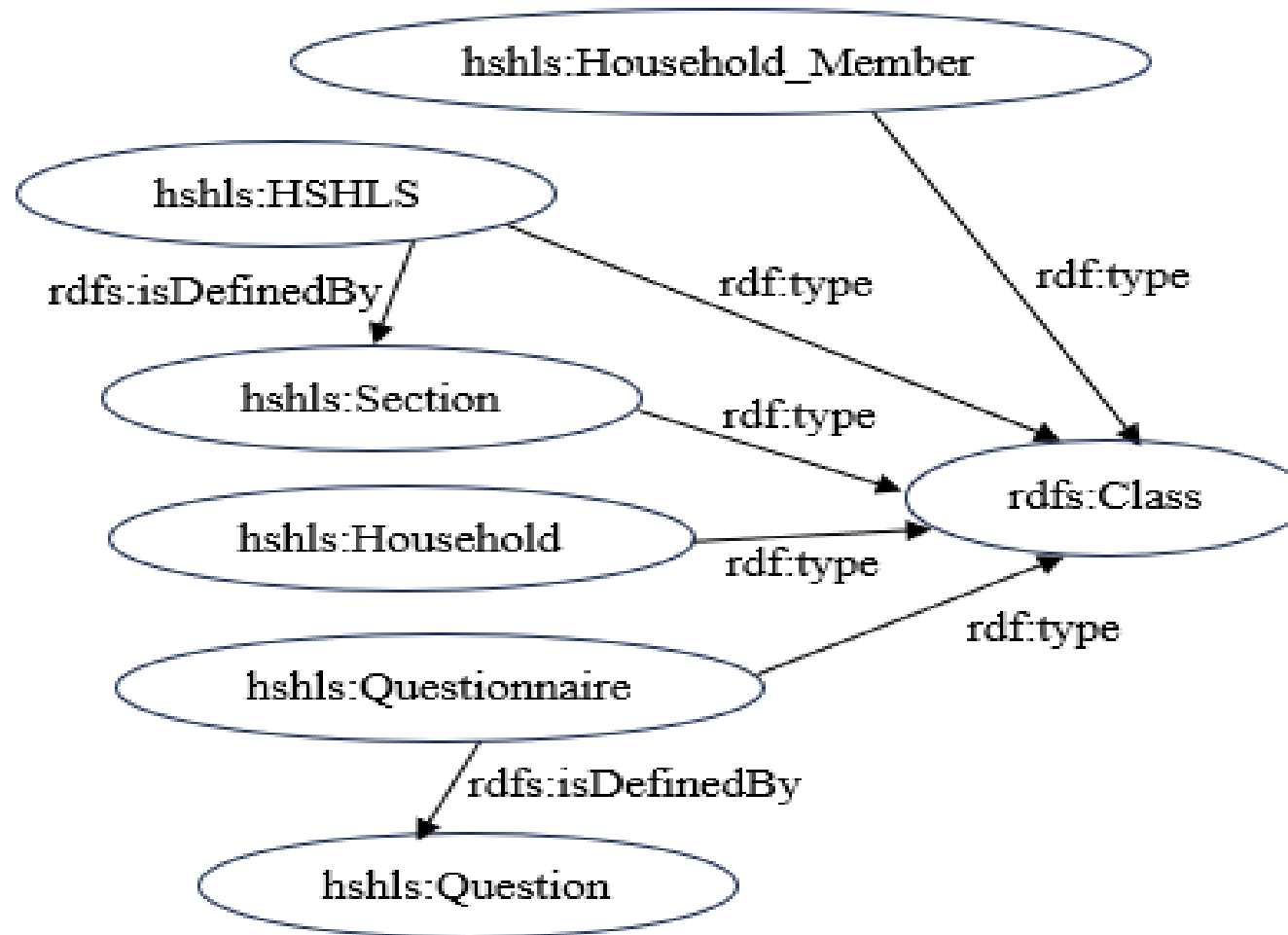
## 2. Aims and contributions

In our paper, we aimed at documenting and disseminating, in a human and computer-readable format, the knowledge contained in the methodological documents of the HSHLS survey.

We designed and implemented a semantic data model of the survey, based on an ontology built using Resource Description Framework (RDF) [2] and its extension, RDF Schema[3].

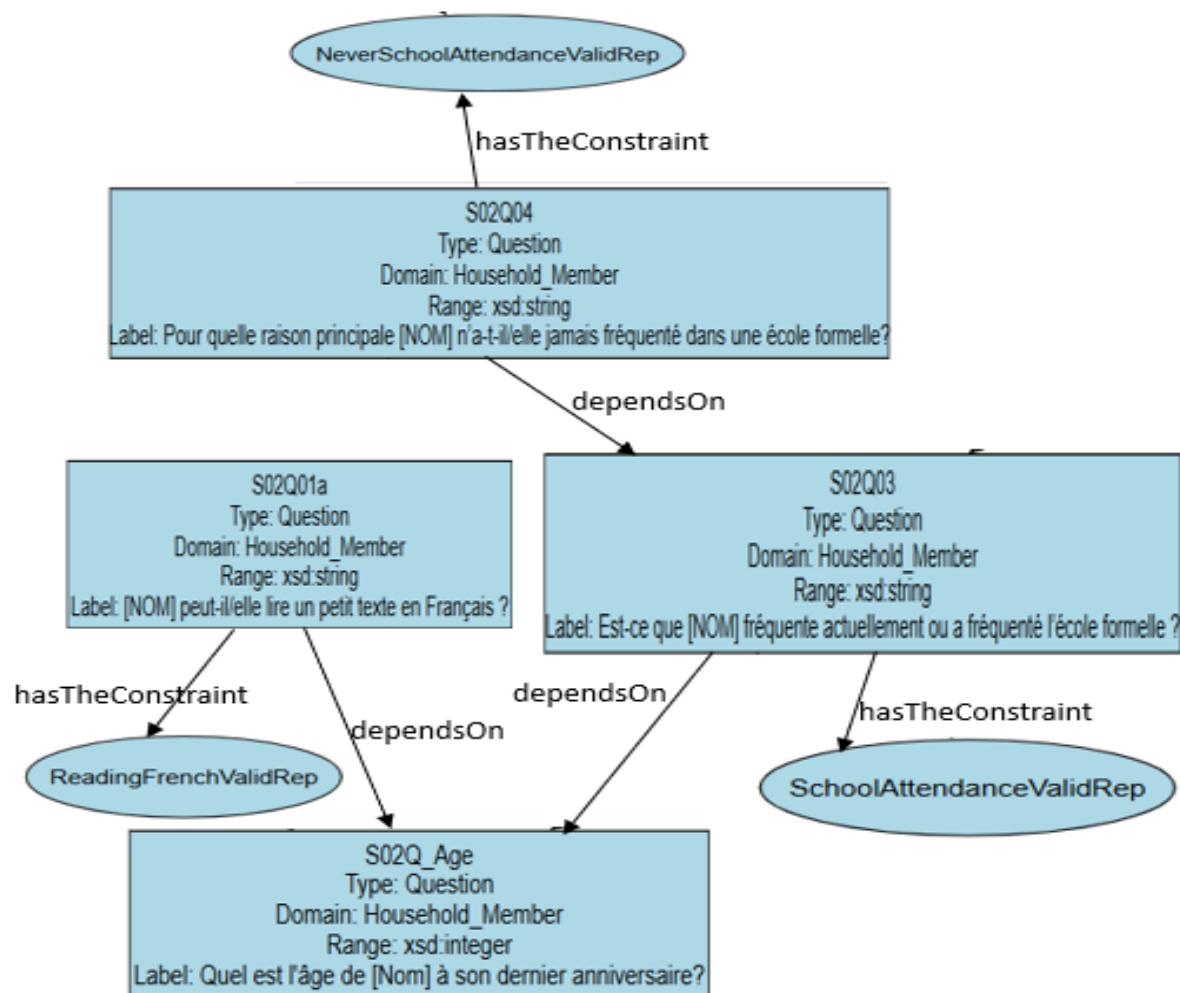
That contributes in enhancing the efficiency of that survey.

## 3.1 HSHLS RDFS metamodel with class declaration

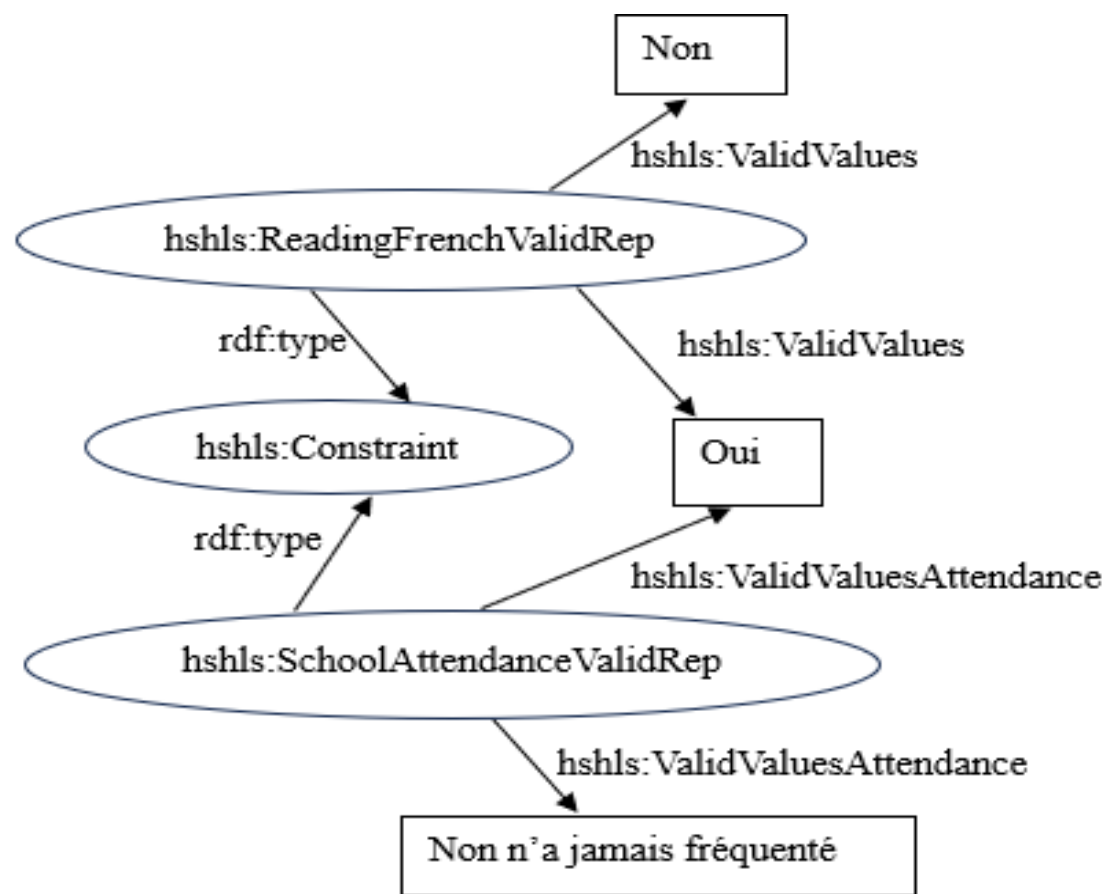




## 4.1 HSHLS RDFS model specialization with actual questions

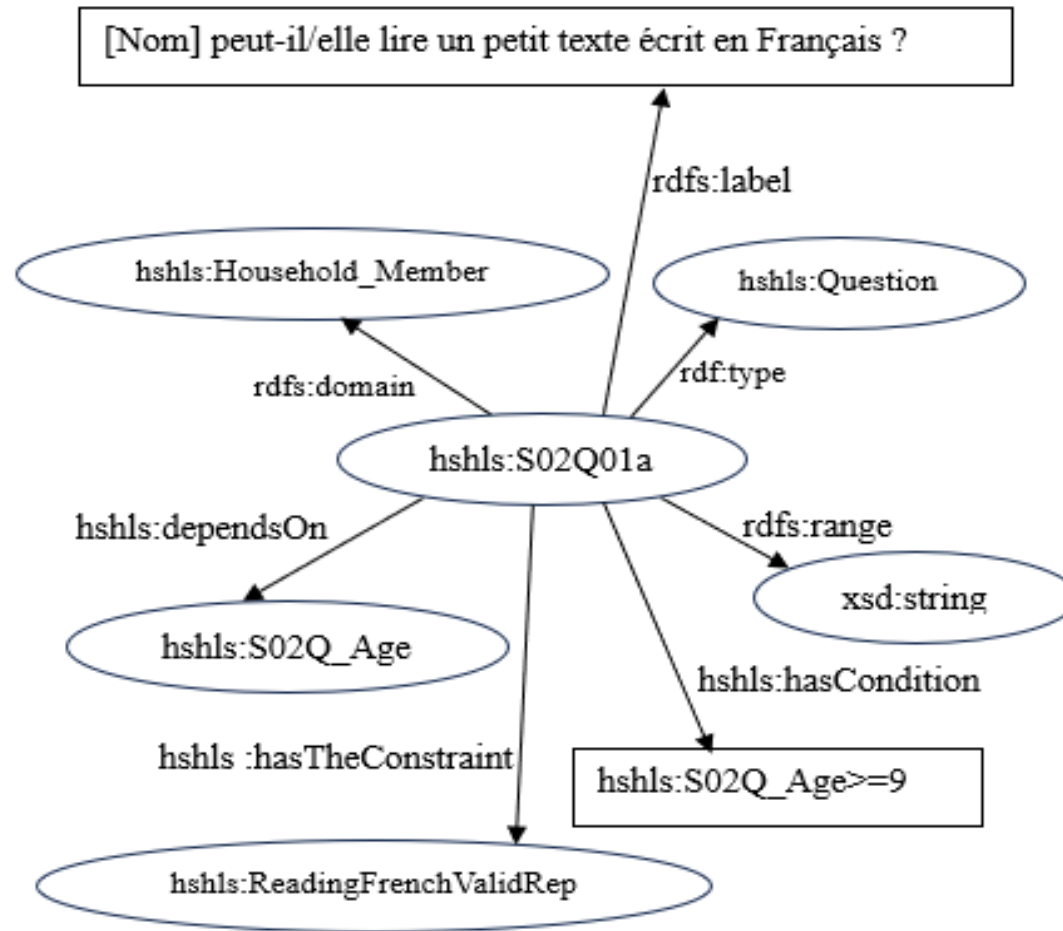


## 4.2 HSHLS RDFS model specialization with actual constraints





## 4.3 HSHLS RDFS model specialization with dependency conditions



## 5. Implementation

We built the proposed model using the Python RDFLib package [4], in a Jupyter notebook environment. The model is saved in Turtle (ttl) format and can be exploited as an RDF graph.

To make it possible to get access to the model in a persistent way, we defined an International Resource Identifier (IRI) for the model.

We also developed an HTML ontology documentation file using PyLODE [5]. We created a public Github repository ([6]) and saved the rdf graph and its HTML documentation in.

## 6. Results and discussion

The HSHLS RDFS semantic model makes it possible to store the methodological information of the HSHLS survey in a human and computer-readable format and retrievable automatically.

By specifying the semantics of the questions addressed, this model helps to better understand the meaning of the data manipulated in this survey as well as the semantic relationships that exist between these data.

With the specification of some constraints and conditions on related questions, this model can serve as a fundamental tool for data quality control on actual data during data collection and processing.

## 5. Results and discussion

Since the model is saved in a persistent repository, one can easily get access and perform some retrievals and analysis requests using SPARQL [7] or any appropriate data analysis tool.

Also, a large audience can get access and learn related knowledge.

Therefore, the project will not only help improving the efficiency during the survey data collection and processing activities, but also contributes to the dissemination of the survey knowledge.

## 6. Conclusion and future work

The HSHLS RDFS semantic model highlights semantic information derived from the methodology of the Harmonized Survey on Households Living Standards. The results of this work can be exploited as part of the automatic retrieval of methodological information.

Generally speaking, this work completes the state of the art and serves as a proof of concept to demonstrate the feasibility of documenting the knowledge contained in a statistical survey questionnaire through ontology-based semantic modeling.

Researchers from a variety of backgrounds will find it a source of information when it comes to design approaches requiring ontology-based semantic modeling of data.

## 6. Conclusion and future work

To enable automatic data quality control based on this model, an extension of the model will be developed in the future, with complex constraints and conditions, using OWL2 [8] or another equivalent language that we will study, which is complementary to and interoperable with RDF.

We will therefore build and implement an automatic reasoning engine to enable anomaly detection in actual data.

# References

- [1] Harmonized Survey on Households Living Standards 2018-2019, Food and Agriculture Organization of the United Nations, 2022, <https://microdata.fao.org/index.php/catalog/2355/study-description>.
- [2] RDF 1.1 Primer, W3C Working Group Note 24 June 2014 <https://www.w3.org/TR/rdf11-primer/>.
- [3] RDF Schema 1.1, W3C Recommendation 25 February 2014, (Document updated on December, 1th 2023), <https://www.w3.org/TR/rdf-schema/>.
- [4] RDFLib, <https://pypi.org/project/rdflib/>.
- [5] PyLODE, <https://pypi.org/project/pylode/>.
- [6] M. Mfoutou Moukala, HSHLS survey ontology Web Page, <https://moukmarc.github.io/>.
- [7] SPARQL 1.1 Query Language, W3C Recommendation 21 March 2013, <https://www.w3.org/TR/sparql11-query/>.
- [8] OWL2 Web Ontology Language, W3C Recommendation 11 December 2012, Document Overview (Second Edition), <https://www.w3.org/TR/owl2-overview/>.