Carlos Utrilla Guerrero

📞 +34 629 95 44 70 ● ✉ c.utrilla.guerrero@gmail.com ● 🖥 carlosug.github.io

Relevant research projects contribution

Distributed Knowledge Graphs for food (KGFOOD): The project started in *Aug 2022*, and aims at building a knowledge graphs on Food Composition Open Datasets and food price sources for identifying sustainable food choices using Knowledge Graphs Embeddings (KGE) as part of EU COST Action on Distributed Knowledge Graphs. ROLE: Daily coordinator of the project, leading data conversion to RDF, generation of standard APIs to query the converted datasets and usecase analysis (e.g. link prediction, completing KG with embedding techniques or network analysis) in *'Food and Health Domain'*. It is expected to submit a conference paper at SWAT4HCLS -Semantic Web and FAIR data in agriculture, food and nutrition domain. *https://github.com/MaastrichtU-IDS/kg-food*

- Semantic web: RDF, SPARQL, Ontologies, RDFLib, Protege
- Languages: Python and Bash
- Research line: Ontology engineering, Knowledge publication and Knowledge graph completion

FAIR Data for traditional games (PLAYFAIR): Since *July 2021*, I collaborated with the Digital Ludeme Project (DLP) team through a joint research initiative called the PLAYFAIR project. This project focused on implementing semantic web in digital humanities, had a *budget of 15,000 euros*, and ended in *March 2022*. ROLE: I took full initiative of the project, from its original formulation, development of the proposal and its execution. Presented as demo at the Conference Annual DARIAH 2022: Storytelling. *https://github.com/MaastrichtU-IDS/play-fair*

- Semantic web: RDF, SPARQL, Ontologies, RDFLib
- Languages: Python and Bash
- Tools: TriplyDB based on CLARIAH infrastructure
- Research line: Semantic interoperability, Knowledge publication and Network science

LD Wizard for Humanities and Social sciences (LDWizard): This small contribution built upon the *PLAYFAIR* project (above) to technically evaluate CLARIAH tools. As a result, we proposed an alternative and built an easy-to-use software to convert tabular data into Linked Data. ROLE: I co-developed the tool and inserted OWL ontologies that are relevant to digital humanities domain. *https://humanities.wizard.semanticscience.org/*

- Semantic web: RDF, SPARQL, Ontologies, RDFLib
- Languages: Python, TypeScript and Bash
- Web dev: HTML/CSS and React
- Research line: Knowledge capture, mappings and metadata editor tools

Data Sharing Coalition (DSC): DSC is a large interdisciplinary research project involving several industry partners. The goal is to promote data sharing practices that are inputs for the development of generic data sharing standards. ROLE: I was particularly involved in building a knowledge base from texts about human trafficking data partnered with Sustainable Rescue NGO and Rosaka. Check an interview in DSC social media about expected contribution to DSC.

- Research line: Cross-domain interoperability, NLP and data sharing infrastructure

Community Data Driven Insights (CDDI): This project consisted of a diverse research project (*"Showcases"*) from Maastricht University, exploring the technical, ethical and social complexity of implementing FAIR. ROLE: I developed ETL processes and workflows for data analysis, and data visualisation, as well as generate conceptual data models and standards for some showcases *'FAIRifing showcases'*, which focuses on the interoperability by annotating the data with ontologies and standard vocabularies. This work has been accepted for the Fair Digital Object (FDO) International Conference in Oct 2022. Conf. article: *https://doi.org/10.3897/rio.8.e96082*

- Semantic web: RDF and Ontologies
- Languages: Python and R
- Database: SPARQL, MongoDB, SQL
- Research line: Cross-domain interoperability, knowledge validation and Open Science

Other projects contribution

Semantics:

- KG for Open Government Data (ESGREEN) – I helped to generate a RDF knowledge graph for green spaces infrastructure data from cities (Python, RML Mappings). *https://github.com/carlosug/esgreen-kg*. A master student was going to further explore the feasibility of using SIO Ontology towards interoperability in a research project started in Oct 2022.
- IDS Project Dashboard – I co-developed a web service to insert projects information from git repositories, alongside with generating JSON file using DOAP schema (Python, React/TypeScript). *https://github.com/MaastrichtU-IDS/projects*
- INDEX – I defined the user requirements for the system developed by Vincent Emonet with a web service to index standard knowledge representations, such as OWL ontologies and SHACL shapes generating metadata for active SPARQL endpoints (Python, React/TypeScript). *http://index.semanticscience.org/*
- A master student research project about Mercedes Benz Claims Automatic Recommendation system powered by semantic technologies (CAR) – (Python, React/TypeScript). *http://index.semanticscience.org/*

Open Science, FAIR and community standards:

- FAIR Enough – I helped define the metrics for FAIR evaluation tool of resources online (Python,React/TypeScript). *https://github.com/MaastrichtU-IDS/fair-enough*
- European Open Science Cloud-Life (EOSC-Life) - I reported on the user requirements and personas for the *'FAIRAssist'* tool that was designed for EOSC-Life actors and profile educational resources (e.g. Generate an ontology for digital skills and competences for EOSC Personas). Check slides
- FAIRness evaluation reports for the Netherlands Consortium of Dementia Cohorts (NCDC) - I utilized FAIR Maturity Evaluation services for evaluating different online resources within NCDC and suggested future directions to ensure resources are compliant with FAIR principles.

Quantitative methods and statistics:

- Developing distributed analytics and privacy preserving techniques in social-emotions domain (ODISSEI) – With colleagues from Inspectorate Directorate, I helped to design a library to generate synthetic tabular or RDF data using Conditional Generative Adversary Networks (GANs) (Python). *https://odissei-data.nl/en/privacy-preserving-techniques/*
- Estimating Time-Varying Models in High-Dimensional EMA Data – Co-supervision of bachelor student Julian Adam that focused on data mapping variables and data analysis using Generalized Additive Modeling (GAM) framework (Python).

Industry and user/department support:

- Data Science Research Infrastructure (DSRI) – I provided technical support to researchers on the HPC cluster, and maintained project website and documentation. I accidentally started to learn Kubernetes for cluster management system and containerization. *https://dsri.maastrichtuniversity.nl/*
- Analysis Ready Data (KLEOS.space) – I resolved issues on data fusion and conducted statistical analysis of different countries with different geolocating datasets using geo standards and R for the European Agency Space
- Data governance and Financial Knowledge model (KNEIP Communications) – I helped resolve data collection and data integration issues in different projects related to EU regulations using ORACLE products. I played with FuzzyWuzzy python package (*https://pypi.org/project/fuzzywuzzy/*) to create a tool for financial term similarities and substitutions.