

Contribution ID: 40

Type: Talk (15min + 5min)

The SPARQL Unicorn: A Research Tool for Linked Open Data in QGIS and git-action-based ontology documentation

Thursday, March 7, 2024 11:00 AM (20 minutes)

Introduction

Publishing sustainable research data and providing appropriate access for many research communities challenges many players: Researchers, RSEs, standardisation organizations and data repositories. With national research data infrastructures (NFDI) being set up in Germany, the latter could be solved in the mid- to longterm for specific datasets. In the meantime, researchers often produce datasets of data in research projects which are provided as services, e.g. from a web page, but may, due to a lack of funding, disappear in that form after the research project has ended. To circumvent this, open research data is hosted long-term on public platforms like university libraries, Zenodo or Github. However, this hosted data is not necessarily easily discoverable by different research communities. On top of that, research data is rarely published in isolation, but with links to related datasets, leading to the creation of link-preserving, FAIR linked open data (LOD) as RDF dumps, modelling data interoperably in common vocabularies. LOD in RDF preserves links, but is not necessarily Linked Open Usable Data (LOUD), i.e. it does not provide data in ways different research communities expect. We would like to address this problem of missing LOUD data while removing requirements on the backend such as hardware and software to a minimum.

Documentation-Tool

We believe that a solution to this data provision problem is publishing research data as static webpages and using standardised static APIs to serve data in ways different research communities expect.

We developed a documentation extension to our SPARQLing Unicorn QGIS Plugin, allowing to publish RDF data dumps as HTML page and RDF serialization per data instance, similar to what frontends to triple stores such as Pubby provide.

It is published as a QGIS Plugin, a standalone script on Github and a Github Action.

The resulting data dump is hostable on static webspaces e.g. Github pages and allows navigating the contents of the LOD data in HTML including a class tree. It may include:

- * Further data formats: Graph Data (GraphML, GEXF), General Purpose (CSV)
- * SPARQL querying in JavaScript using the data dump
- * Generation of static APIs, e.g. JSON documents mimicking standardized APIs, for
- * OGC API Features: Access to FeatureCollections from e.g. QGIS

* IIIF Presentation API 3.0: IIIF Manifest Files for images/media in the knowledge graph including typed collections

* CKAN API: Datasets in the DCAT vocabulary or data collections

Static APIs further the accessibility of LOD data for different research communities and increase the chances of data reusage and exposure in different research fields, while at the same time not depending on additional infrastructures for data provision.

Limitations and Future Work

Our talk shows the feasibility of using publicly available examples for geodata and CKAN (SPP Dataset, AncientPorts Dataset, CIGS Dataset) and the ARS-LOD dataset for static IIIF-data.

We discuss requirements and limitations of this kind of publishing in a RDM publishing workflow, in relation to NFDI plans and how to extend this approach to only partially open data using a Solid pod publishing workflow.

Slot length

Primary authors: THIERY, Florian (Leibniz-Zentrum für Archäologie (LEIZA)); HOMBURG, Timo (Hochschule Mainz)

Presenters: THIERY, Florian (Leibniz-Zentrum für Archäologie (LEIZA)); HOMBURG, Timo (Hochschule Mainz)

Session Classification: Metadata for Research Software

Track Classification: Research Software: FAIRification and Its Implications for Research Software