



Contribution ID: 108

Type: **Talk (15min + 5min)**

Jupyter Python Minion: Simplifying SPARQL Queries and Visualisations for Archaeological Data

Thursday 27 February 2025 11:20 (20 minutes)

The Semantic Web is a treasure trove of interconnected knowledge graphs, providing access to datasets that are invaluable for research in cultural heritage and archaeology. Resources such as triplestores (e.g., the NFDI4Objects Knowledge Graph), Wikibase instances (e.g., Wikidata and FactGrid), and Solid Pods housing geoscientific data open new avenues for interdisciplinary exploration. However, researchers face significant challenges in utilising these resources effectively. Writing SPARQL queries requires a steep learning curve, and the data often returns in formats like `sparql-results+xml` or `sparql-results+json`, which are not user-friendly for immediate analysis or visualisation.

Python has become a critical tool in addressing these challenges. As a versatile scripting language, Python enables researchers to automate workflows, ensure reproducibility, and integrate datasets seamlessly. However, many researchers lack the technical skills or frameworks needed to implement Python solutions in their work. This is where Jupyter Notebooks provide a critical advantage. Combining an intuitive, shareable environment with the computational power of Python, Jupyter Notebooks make it easy to share not just results but the entire workflow. This transparency enhances reproducibility, facilitates collaboration, and aligns with FAIR principles.

The **Jupyter Python Minion** builds on this framework, offering an open-source solution to simplify SPARQL querying and data visualisation. By integrating widely used Python libraries such as `pyplot`, `wordcloud`, `geopandas`, and `contextily`, it transforms raw Linked Open Data (LOD) into actionable insights. Researchers can produce bar charts, pie charts, maps, and word clouds with minimal effort, bridging the gap between technical expertise and domain-specific knowledge. Importantly, the tool enables researchers to document their computational workflows within Jupyter Notebooks, creating reusable resources for the broader research community.

The need for such tools in archaeological research software engineering is pressing. Computational archaeology increasingly relies on integrating diverse datasets—geospatial, semantic, and cultural—but the technical barriers often hinder broader adoption. By lowering these barriers, the Jupyter Python Minion empowers researchers to harness the power of Python scripts for reproducible and shareable analysis.

In this talk, I will demonstrate the utility of the Jupyter Python Minion through five use cases:

1. **Playful exploration of Pokémon properties**, demonstrating SPARQL queries and visualisation techniques like scatterplots and bar charts.
2. **Mapping the distribution of Samian ware kiln sites**, showcasing the regional breakdown of archaeological production centres.
3. **Exploring Irish Holy Wells**, revealing etymological patterns through word clouds and pie charts.
4. **Geospatial analysis of Irish Ogham Stones**, including density maps and OpenStreetMap-based visualisations to highlight regional clusters.
5. **Integration of geoscientific findspots from Solid Pods**, using SPARQL queries to categorise archaeological and geological locations affected by the Campanian Ignimbrite eruption.

These examples highlight the transformative potential of integrating Python scripting with Jupyter Notebooks for reproducible research. The tool's shareability fosters collaboration across disciplines, from archaeology to geosciences, and promotes a culture of openness and accessibility in research software engineering.

This talk will contribute to key themes in RSE, including computational workflows, open-source tools, and software usability, while providing attendees with actionable insights to adopt and adapt these methods in

their own research.

I want to participate in the youngRSE prize

no

Primary authors: THIERY, Florian (Research Squirrel Engineers Network); SCHUBERT, Lutz Krister (University of Cologne)

Presenters: THIERY, Florian (Research Squirrel Engineers Network); SCHUBERT, Lutz Krister (University of Cologne)

Session Classification: Visualization with Research Software

Track Classification: Research Software: domain-specific languages