

The FAIR Signposting Crawler

Building Knowledge Graphs from FAIR Digital Objects on the Web

Thivisha Uthayakumar¹, and Stian Soiland-Reyes¹ 

¹Department of Computer Science, The University of Manchester, UK

*Correspondence: Stian Soiland-Reyes, soiland-reyes@manchester.ac.uk

Abstract. Despite Linked Data being an established a framework for structured meta-data, significant barriers remain in automatically discovering and accessing these resources across the web. The premise of FAIR Digital Objects (FDO) adds a navigational overlay with predictable semantics for consistent machine-actionability. FAIR Signposting is recognised as part of a Web-based implementation of the FDO principles, using typed links from HTTP headers and HTML landing pages to constituent FDO elements, including persistent identifiers and metadata resources using Linked Data. While FDOs by design are inherently interconnected, currently no general consumers of a network of “Webby FDOs” have been implemented. Here we present the FAIR Signposting Crawler to demonstrate systematic FDO discovery and knowledge graph construction, enabling automated traversal of interconnected Web resources without requiring human intervention.

Keywords: FAIR Digital Objects, Signposting, Knowledge Graphs, Provenance, FAIR Assessment, Linked Data, SPARQL

Introduction

The widely endorsed FAIR principles detail how research data should be made Findable, Accessible, Interoperable, and Reusable, and have largely been implemented using metadata in the form of Linked Data, with persistent identifiers to machine-readable resources that either mirrors corresponding human-readable resources, or present both using HTTP-level content negotiation based on file formats. However, with a wide range of metadata vocabularies, and multiple RDF data formats and PID deployment scenarios, consistent machine-based processing necessitates further navigational formalisation, as suggested by the FAIR Digital Object (FDO) specifications. FAIR Signposting, based on embedded typed links in HTTP headers and HTML documents, has been combined with the research data packaging method Research Object Crate (RO-Crate) to form the “Webby FDO”, now recognised as a implementation of FDO [1].

Despite their potential, tools for consuming Signposting links have been limited to rendering and parsing of individual FDOs. This work explores how Signposting can be used to construct knowledge graphs from distributed FDOs. Rather than attempting

to implement a full Webby FDO client, we focus on demonstrating how Signposting and linked metadata can be used to discover and integrate resources into a coherent RDF-based knowledge graph. This approach helps to illustrate the advantage of linked FDOs, as the graph can be queried and visualised in powerful ways.

Implementation

Signposting Crawler has a modular architecture with a configurable crawler engine that identifies Signposting links in HTTP headers and HTML elements, while employing adaptive fallback methods when direct Signposting is unavailable. The implementation includes an RDF processing pipeline that transforms discovered data into a coherent knowledge graph stored in Apache Jena Fuseki, and a FAIR Assessment module that evaluates resource compliance.

Developed using Python and Flask, the web interface provides interactive visualisations, a SPARQL query interface, and comprehensive assessment reporting. This work demonstrates how Signposting can effectively reduce barriers between traditional web resources and their machine-actionable counterparts, advancing the broader FDO goal of truly machine-actionable research data on the web.

Type of Contribution

This submission is for presenting a **Poster**.

Underlying and related material

Source code for the Signposting Crawler is archived from GitHub repository [thivi19/third-year-project](#) is archived at Zenodo with DOI <https://doi.org/10.5281/zenodo.15059489>

Author contributions

TU: Conceptualisation, Data Curation, Investigation, Software, Visualization
SSR: Conceptualisation, Supervision, Writing – original draft

This abstract is based on a report with the same name submitted by TU to The University of Manchester for the degree of Bachelor of Science.

Competing interests

The authors declare that they have no competing interests.

References

- [1] S. Soiland-Reyes, P. Sefton, S. Leo, L. J. Castro, C. Weiland, and H. Van de Sompel, "Practical webby fdos with ro-crate and fair signposting: Experiences and lessons learned," *Open Conference Proceedings*, vol. 5, Mar. 2025. DOI: [10.52825/ocp.v5i.1273](https://doi.org/10.52825/ocp.v5i.1273).