

Appendix: Content Summaries of Selected Best Papers Published in 2022 for the IMIA Yearbook, Section Knowledge Representation and Management

Kaliyaperumal R, Wilkinson MD, Moreno PA, Benis N, Cornet R, dos Santos Vieira B, Dumontier M, Bernabé CH, Jacobsen A, Le Cornec C, Godoy MP

Semantic modelling of common data elements for rare disease registries, and a prototype workflow for their deployment over registry data

J Biomed Semantics 2022;13(1):9. doi: 10.1186/s13326-022-00264-6

In this article, the members of the EU Platform on Rare Disease Registration put forward a working procedure to ensure the interoperability and FAIRification of data from the common data representation model, the CDE. The motivation is precise and industrial: the 16 CDEs must be implemented in all EU Rare Disease registries. To do this, they created semantically grounded models to represent each of the CDEs, using the SemanticScience Integrated Ontology as the core framework for representing the entities and their relationships. Within

that framework, they mapped the concepts represented in the CDEs, and their possible values, to domain ontologies such as the Orphanet Rare Disease Ontology, Human Phenotype Ontology and National Cancer Institute Thesaurus. Finally, they created an exemplar, reusable ETL pipeline that they will be deploying over non-coordinating data repositories to assist them in creating model-compliant FAIR data without requiring site-specific coding, nor expertise in Linked Data or FAIR. This ETL refers to alignment description languages and execution models with YAML and YARRRML. With the aforementioned ontologies, the authors describe an industrial process at the knowledge level, perfectly operational, and in total respect of FAIR principles.

Matentzoglu N, Goutte-Gattat D, Tan SZ, Balhoff JP, Carbon S, Caron AR, Duncan WD, Flack JE, Haendel M, Harris NL, Hogan WR

Ontology Development Kit: a toolkit for building, maintaining and standardizing biomedical ontologies

Database 2022:baac087. doi: 10.1093/database/baac087

In this paper, the authors provided an overview of the Ontology Development Kit (ODK), a Docker-based tool for creating and managing ontologies in the biomedical domain. ODK consists of a toolbox equipped with diverse tools for ontology editors to

build, test, and release ontologies as well as a set of standardized and executable ontology-engineering workflows following the best practices recommended by the Open Biological and Biomedical Ontology (OBO) Foundry. Moreover, the authors attempt to highlight how ODK stimulates standardization efforts in the Knowledge Representation (KR) community. The authors have already observed significantly lower error rates in many of the ontologies that use the ODK, thanks to the ability of the automated testing system provided by the ODK to catch errors early on. Lastly, they seek to harmonize the representation of ontology release files through the use of standard release workflows, which result in standard release serializations and metadata to make ontologies more FAIR and interoperable. ODK is not an ontology editor. It lets ontology creators use an editor like Protégé. ODK supports a templating system such as ROBOT or others. This is not the first time that researchers in the KR domain have proposed an ontology development environment: NEON Toolkit or WebODE are proposals that have been around for over 10 years. ODK is more recent, takes into account modular approaches that are now important, and respects the de facto standards of ontology engineering. Time will tell whether it will become the reference tool. In any case, we need ontology development methodologies and softwares to implement them.