Meet PUDEL – A New Service for Sharing and Documenting Data Models

Becker, Anja
becker@saw-leipzig.de
Saxon Academy of Sciences and Humanities, Germany

Graiff, Cecilia
graiff@saw-leipzig.de
Saxon Academy of Sciences and Humanities, Germany

Goldhahn, Dirk
goldhahn@saw-leipzig.de
Saxon Academy of Sciences and Humanities, Germany

Kretschmer, Uwe
kretschmer@saw-leipzig.de
Saxon Academy of Sciences and Humanities, Germany

Mühleder, Peter
muehleder@saw-leipzig.de
Saxon Academy of Sciences and Humanities, Germany

Naether, Franziska
naether@saw-leipzig.de
Saxon Academy of Sciences and Humanities, Germany

Creating ‘data models’ is a central activity in the Digital Humanities. On the most fundamental level, the term refers to "the modeling of some segment of the world in such a way to make some aspects computable, referring to creating database schemas, SGML DTDs, XML schemas, ontologies etc." (Flanders / Jannidis 2015). But, despite being an important step in many DH workflows, the documentation, publication and reusability of these data models often does not play a central role in the general research process, making the conceptual work behind the models difficult to find, access and comprehend. Our goal is therefore to create a service that allows researchers and DH specialists an easy way of documenting, disseminating and reusing their data models: PUDEL, the Publication Service for Academic Data Models and Vocabularies. The project is funded by the Initiative for Research Data Management in Saxony (SaxFDM) and developed at the Saxon Academy of Sciences and Humanities in Leipzig (Germany).

Existing services, like the Vocabs Service of ACDH-CH (Austrian Academy of Sciences 2023), only support the publication of data models in SKOS format, or require the data model to already be published online to be able to reference it, as it is the case with VoCoReg (Fraunhofer Society 2023). PUDEL attempts to provide a more general solution, avoiding these limitations and allowing for a more comprehensive approach for sharing data models. Based on an extended review of existing tools and by creating user stories based on real world use cases, we outlined three basic pillars for the service: simplicity, transparency and openness. Each of these concepts are reflected both in the usability as well as in the technical design of the service. Fig. 1 summarizes the main features of PUDEL.

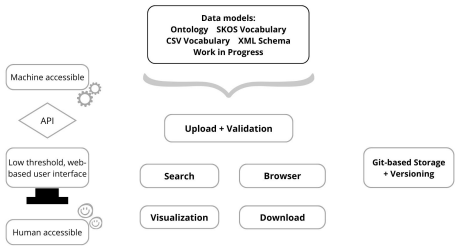

Fig. 1: Simplified scheme of the main features of the new publishing service for data models PUDEL.

(1) Simplicity: The goal of PUDEL is to provide a low barrier workflow to publish and search for data models. For this reason, an intuitive web interface is created as an entry point for the service and various ‘middleware’ services validate data models and create representations based on established best practices (Garijo / Poveda-Villalón 2020; Semantic Web Deployment Working Group 2008). Additionally, PUDEL supports automatic publication on the open repository Zenodo (European Organization For Nuclear Research, OpenAIRE 2013). Besides user experience, PUDEL, as a set of software tools, should also be ‘simple’ to maintain and update. Here, PUDEL follows the approach of minimal computing (Sayers 2016) whenever possible, using few resources and avoiding a complex system architecture.

(2) Transparency: An important aspect of PUDEL is the possibility to not only publish, but also document the evolution of data models. PUDEL uses a Git-based file storage backend, which allows users to update their data models without losing access to previous versions. This versioning is also part of the stable identifiers provided by PUDEL, allowing referencing a specific version of an ontology, class or concept.

(3) Openness: PUDEL is oriented at the FAIR-principles for research data: findability, accessibility, interoperability, and reusability. Alongside with supporting the reuse by individuals, these principles facilitate machines in automatically finding and using the data (Wilkinson et al. 2016). As a service, PUDEL embraces openness by supporting multiple widely used data formats for data models (RDF based ontologies, XML schemas) and vocabularies (SKOS, CSV-based lists). In order to allow easy access for external services and tools, the OpenAPI standard (OpenAPI Initiative 2022) is used to create comprehensive and standardized documentations for the service.

PUDEL is developed as a free and open-source software (FOSS), making it possible for institutions to run their own instance of PUDEL and allowing them to publish data models within their own namespaces. Altogether, by following the above principles, PUDEL is a service that can easily be included into existing DH research practices and provides institutions, researchers and specialists with the tools to sharing and preserving their data models.

Bibliography

Austrian Academy of Sciences , Austrian Centre for Digital Humanities and Cultural Heritage (ACDH-CH), "Vocabs services", https://vocabs.dariah.eu/ [25.04.2023].

European Organization For Nuclear Research, OpenAIRE (2013): "Zenodo", in: CERN Publ., doi: 10.25495/7GXK-RD71, https://www.zenodo.org/ [25.04.2023].

Flanders, Julia / **Jannidis, Fotis** (2015): "Knowledge Organization and Data Modeling in the Humanities", https://nbn-resolving.org/html/urn:nbn:de:bvb:20-opus-111270 [25.04.2023].

Fraunhofer Society , "VoCol Service on VoCoREG", https://www.vocoreg.com/ [25.04.2023]. **Garijo, Daniel** / **Poveda-Villalón, María** (2020): "Best Practices for Implementing FAIR Vocabularies and Ontologies on the Web", https://doi.org/10.48550/arXiv.2003.13084 [25.04.2023].

OpenAPI Initiative (2022): "OpenAPI Specification v3.1.0", in: The Linux Foundation (Publ.), https://spec.openapis.org/oas/latest.html [25.04.2023]

Sayers, Jentery (2016): "Minimal Definitions", in: Minimal Computing Working Group of GO::DH (Publ.), https://go-dh.github.io/mincomp/thoughts/2016/10/02/minimal-definitions/ [25.04.2023].

Semantic Web Deployment Working Group (2008): "Best Practice Recipes for Publishing RDF Vocabularies" (Work in Progress), https://www.w3.org/TR/swbp-vocab-pub/ [25.04.2023].

Wilkinson, Mark D. et al. (2016): "The FAIR Guiding Principles for scientific data management and stewardship", in: Sci Data 3: 160018, https://doi.org/10.1038/sdata.2016.18 [25.04.2023].