

# The Research Core Dataset (KDSF) in Linked Data context

## Introduction

In this paper, the adaption of KDSF for its implementation in VIVO is described.

The Research Core Dataset (Kerndatensatz Forschung, KDSF) is a new standard for collecting and reporting research information. It is meant to support research institutions in their reporting efforts.

The standard delivers a set of categories for uniform data recording and aggregating. The set includes seven superior categories for recording: person, doctoral programme, third-party funded project, publication, patent, and research infrastructure. There is an aggregational level where there are the following core and peripheral elements: employees, promotion, third-party funds, patents, publications, and research infrastructure. (Institut für Forschungsinformation und Qualitätssicherung (iFQ); Fraunhofer-Institut für Angewandte Informationstechnik; Geschäftsstelle des Wissenschaftsrates, p. 5ff)

The standard is currently being implemented across German academic and research-related institutions. Since the integration of the elements of KDSF into a research information system (CRIS) is regarded as a viable option, several efforts to implement the standard in different types of CRIS like Pure, Converis or various in-house developments were undertaken.

KDSF is available in Web Ontology Language (OWL) format ([http://www.kerndatensatz-forschung.de/version1/technisches\\_datenmodell/owl.html](http://www.kerndatensatz-forschung.de/version1/technisches_datenmodell/owl.html)). It is therefore suitable to be implemented within a Linked Data application.

KDSF specification mentions to VIVO as an eligible software for its technical implementation. (Institut für Forschungsinformation und Qualitätssicherung (iFQ); Fraunhofer-Institut für Angewandte Informationstechnik; Geschäftsstelle des Wissenschaftsrates, p. 24ff)

VIVO is an application based on Linked Data technologies. It aggregates research related information, interlinks it semantically and represents it accordingly to web standards. As an open source software, VIVO is flexible and can be easily adapted to different needs. As a Linked Data application, VIVO is built upon on a knowledge base that models information using established Linked Data vocabularies such as Friend of a Friend Ontology (FOAF), Bibliographic Ontology (BIBO), Geopolitical Ontology, Simple Knowledge Organization System (SKOS) and others. (Duraspace)

## Motivation

Implementing KDSF in VIVO enables the use of research information in different contexts. On the one hand, information about people, organizations, projects and publications in form of profiles can be used for representation on the web. The integration of a VIVO compatible ontology developed from the base data model (Basisdaten) of KDSF supports data acquisition and reporting centrally within a single application.

Collecting data in VIVO using linked data technologies has several advantages:

1. In VIVO, the information is not merely recorded and used as character strings, but rather as classified, semantically linked, URI-named and reusable items. One-time collection is sufficient to link one entity with another in a variety of ways and to be able to use it again and again.
2. Inference rules are built into the VIVO data model. They are interpreted by the inference engine and allow the implicit knowledge to be derived.
3. Classes and properties already present in VIVO could be re-used to reflect the facts described in KDSF. The data model can be extended at any time and as required.
4. In VIVO, data can be centrally administered, e.g. entered manually by staff members. Semi-automatic and automatic imports of larger amounts of data are also possible. In this way, on the one hand, the number of people who have access to sensitive data can be greatly reduced. On the other hand, the reporting processes run more efficiently.

5. VIVO is an open source application with an active community. It can be operated and adapted locally. There is no vendor lock-in and the data sovereignty remains with the organization itself.
6. A rights and role system, as well as meta properties (annotation properties) in VIVO, allow settings for viewing, publishing and editing data in VIVO to be made according to local preferences and specifications (data protection, etc.).

Therefore, the goal of this project is to integrate KDSF into VIVO to enable KDSF-compliant reporting by means of structured, semantically supported and user-friendly data collection and report generation.

## Implementation

KDSF is available as an OWL ontology. Thus, classes, class hierarchies, relationships and properties already defined in this ontology are at least formally Linked Data compliant. Integration of the KDSF into VIVO required merging of the KDSF base data model and VIVO ontology.

Since for both VIVO ontology and KDSF there are few rules of inference and no complex relationships, e.g., disjunction, functional or inverse relationships, the computer-aided ontology matching was omitted during the merging process.

One method of merging was to identify and apply equivalent relationships between KDSF categories and those of the vocabularies provided in VIVO. Another method was to insert KDSF elements as subsets of VIVO categories.

Although KDSF is available in OWL format, its classes and properties are not necessarily designed for use within a semantic application. For example, some properties currently designed as datatype properties would be more valuable if used for interlinking entities instead of the current design of attributing them with strings.

The processes described above have so far resulted in the first draft of a KDSF-VIVO alignment and a KDSF-VIVO extension. (<https://github.com/VIVO-DE/VIVO-KDSF-Integration>) Both drafts have been successfully tested in VIVO 1.9. The outcomes of the project are going to be applied for reporting at our institution and will be made available for use by the whole German academic and research related sector.

## References

Duraspace. (n.d.). *VIVO*. Retrieved Februar 28, 2018, from About VIVO:

<http://vivoweb.org/info/about-vivo>

Institut für Forschungsinformation und Qualitätssicherung (iFQ); Fraunhofer-Institut für Angewandte Informationstechnik; Geschäftsstelle des Wissenschaftsrates. (n.d.). *Kerndatensatz Forschung : Dokumentation der Ergebnisse ; 13 Oktober 2015*. Retrieved from

[http://www.kerndatensatz-forschung.de/version1/Ergebnisbericht\\_Projekt\\_KDSF.pdf](http://www.kerndatensatz-forschung.de/version1/Ergebnisbericht_Projekt_KDSF.pdf)

Institut für Forschungsinformation und Qualitätssicherung (iFQ); Fraunhofer-Institut für Angewandte Informationstechnologie; Geschäftsstelle des Wissenschaftsrates. (n.d.). *Spezifikation des Kerndatensatz Forschung : Version 1.0 ; 24. September 2015*. Retrieved from

[https://kerndatensatz-forschung.de/version1/Spezifikation\\_KDSF\\_v1.pdf](https://kerndatensatz-forschung.de/version1/Spezifikation_KDSF_v1.pdf)