

ODISSEI Knowledge Graph RDF Schema Documentation

This document provides a detailed overview of the RDF schema for the ODISSEI Knowledge Graph, extracted from its SPARQL endpoint. The schema defines the classes, concepts, and relations used to structure the data within the knowledge graph.

1. Introduction

The ODISSEI (Open Data Infrastructure for Social, Economic and behavioral sciences in the Netherlands) Knowledge Graph aims to integrate and provide access to a vast array of social science data. Understanding its underlying RDF schema is crucial for effective data discovery, querying, and utilization. This documentation is automatically generated by Manus AI based on direct queries to the ODISSEI SPARQL endpoint located at <https://api.kg.odissei.nl/datasets/odissei/odissei-knowledge-graph-acceptance-virtuoso/sparql>.

2. Methodology for Schema Extraction

The RDF schema was extracted by querying the ODISSEI SPARQL endpoint for distinct classes (types) and predicates (properties). The following SPARQL queries were used:

To retrieve distinct predicates:

```
SELECT DISTINCT ?p WHERE { ?s ?p ?o } LIMIT 100
```

To retrieve distinct classes:

```
SELECT DISTINCT ?type WHERE { ?s a ?type } LIMIT 100
```

These queries provided a raw list of URIs for both classes and predicates present in the knowledge graph. To focus on the ODISSEI-specific schema, a filtering process was applied. Common RDF, RDFS, OWL, and SKOS vocabulary terms, as well as general schema.org terms, were excluded. The remaining URIs were then categorized by their namespaces, with particular attention paid to namespaces containing 'odissei' and 'dataverse', as these are

central to the ODISSEI Knowledge Graph. Additionally, relevant external vocabularies frequently used in research and software description were included to provide a more comprehensive view of the domain-specific schema.

3. Schema Overview

The ODISSEI Knowledge Graph schema encompasses a comprehensive set of classes and predicates designed to represent social science research data, datasets, software, and related metadata. The analysis reveals a total of 72 distinct classes and 100 predicates, with 9 domain-specific classes and 65 domain-specific predicates that are particularly relevant to the ODISSEI ecosystem.

3.1 Namespace Distribution

The schema is organized around several key namespaces that reflect different aspects of the knowledge graph:

ODISSEI-specific namespaces	(7 total):
https://portal.odissei.nl/schema/geospatial#	- Geographic and spatial data properties
https://portal.odissei.nl/schema/dansMetadata#	- DANS (Data Archiving and Networked Services) metadata
https://portal.odissei.nl/schema/enrichments#	- Data enrichment and linkage information
https://portal.odissei.nl/schema/provenance#	- Data provenance and workflow tracking
https://portal.odissei.nl/schema/questionInformation#	- Survey question and instrument metadata
https://portal.odissei.nl/schema/socialscience#	- Social science research methodology
https://portal.odissei.nl/schema/variableInformation#	- Variable and data structure information

Dataverse namespaces (2 total): - <https://dataverse.org/schema/citation/> - Citation and bibliographic metadata - <https://dataverse.org/schema/core#> - Core Dataverse functionality

External vocabularies: The schema also incorporates terms from established vocabularies including Schema.org, FOAF, Dublin Core, OWL, SKOS, and specialized research vocabularies such as NWO research fields and software type classifications.

4. Classes (Concepts)

The ODISSEI Knowledge Graph employs a rich set of classes to categorize different types of entities within the social science research ecosystem. These classes range from fundamental RDF/OWL constructs to highly specialized domain-specific concepts.

4.1 Core Research Entity Classes

The schema includes several key classes that represent the primary entities in social science research:

Dataset and Data Resources: - <http://schema.org/Dataset> - Represents research datasets - <http://purl.org/dc/dcmitype/Dataset> - Dublin Core dataset type - <http://schema.org/DataDownload> - Downloadable data resources - <http://schema.org/DataCatalog> - Collections of datasets

Publications and Documentation: - <http://schema.org/ScholarlyArticle> - Academic publications - <http://purl.org/ontology/bibo/AcademicArticle> - Bibliographic academic articles - <http://schema.org/TechArticle> - Technical documentation - <http://schema.org/CreativeWork> - General creative works - <http://schema.org/Book> - Book publications

Software and Technology: - <http://schema.org/SoftwareApplication> - Software applications - <http://schema.org/SoftwareSourceCode> - Source code repositories - <http://schema.org/WebApplication> - Web-based applications - <http://purl.org/dc/dcmitype/Software> - Software resources

4.2 Domain-Specific Classes

The schema includes several specialized classes that are particularly relevant to the ODISSEI research context:

Research Classification: - <https://w3id.org/nwo-research-fields#ResearchField> - NWO (Dutch Research Council) research field classifications, enabling systematic categorization of research projects and datasets according to established Dutch academic standards

Technology Assessment: - <https://w3id.org/research-technology-readiness-levels#TechnologyReadinessLevel> - Technology Readiness Level classifications for assessing the maturity and development stage of research technologies and tools

Software Categorization: The schema incorporates a comprehensive software type taxonomy from the w3id.org vocabulary: - <https://w3id.org/software-types#CommandLineApplication> - Command-line interface tools - <https://w3id.org/software-types#ServerApplication> - Server-based applications - <https://w3id.org/software-types#SoftwareLibrary> - Reusable software libraries - <https://w3id.org/software-types#WebApplication> - Web-based applications - <https://w3id.org/software-types#SoftwareImage> - Containerized software images - <https://w3id.org/software-types#DesktopApplication> - Desktop applications

People and Organizations: - <http://xmlns.com/foaf/0.1/Person> - Individual researchers and contributors - <http://schema.org/Person> - General person entities - <http://schema.org/Organization> - Research institutions and organizations

4.3 Metadata and Structure Classes

The schema includes various classes for organizing and structuring metadata:

Conceptual Organization: - <http://www.w3.org/2004/02/skos/core#Concept> - SKOS concepts for controlled vocabularies - <http://www.w3.org/2004/02/skos/core#ConceptScheme> - Concept schemes for organizing vocabularies

Resource Aggregation: - <http://www.openarchives.org/ore/terms/Aggregation> - Collections of related resources - <http://www.openarchives.org/ore/terms/AggregatedResource> - Individual resources within aggregations - <http://www.openarchives.org/ore/terms/ResourceMap> - Maps describing resource relationships

5. Predicates (Relations)

Predicates define the relationships between entities and their attributes within the ODISSEI Knowledge Graph. They represent the properties that connect subjects and objects in RDF triples. The schema utilizes a mix of standard RDF/RDFS/OWL properties and highly specific domain-oriented predicates.

5.1 Core RDF/RDFS/OWL Predicates

These are fundamental predicates used for defining the structure and semantics of the RDF graph:

- <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> (a) - Indicates the RDF class of a resource.
- <http://www.w3.org/2000/01/rdf-schema#subClassOf> - Defines a subclass relationship between two classes.
- <http://www.w3.org/2000/01/rdf-schema#subPropertyOf> - Defines a subproperty relationship between two properties.
- <http://www.w3.org/2002/07/owl#equivalentClass> - States that two classes have the same extension.
- <http://www.w3.org/2002/07/owl#equivalentProperty> - States that two properties have the same extension.
- <http://www.w3.org/2002/07/owl#inverseOf> - Defines an inverse relationship between two properties.
- <http://www.w3.org/2000/01/rdf-schema#label> - Provides a human-readable name for a resource.
- <http://www.w3.org/2000/01/rdf-schema#comment> - Provides a human-readable description of a resource.
- <http://www.w3.org/2000/01/rdf-schema#domain> - Specifies the class of subjects that a property can have.
- <http://www.w3.org/2000/01/rdf-schema#range> - Specifies the class of objects that a property can have.
- <http://www.w3.org/2000/01/rdf-schema#seeAlso> - Indicates a resource that might provide additional information about the subject resource.

5.2 Dublin Core and Schema.org Predicates

Commonly used predicates for general metadata and descriptive information:

- <http://purl.org/dc/terms/created> - Date of creation of the resource.
- <http://purl.org/dc/terms/modified> - Date of last modification of the resource.
- <http://purl.org/dc/terms/dateSubmitted> - Date of submission of the resource.
- <http://schema.org/additionalType> - An additional type for the item, typically used for adding more specific types from other vocabularies.
- <http://schema.org/creativeWorkStatus> - The status of a creative work (e.g.,

draft, published, etc.). - <http://schema.org/fileFormat> - The file format of the media object. - <https://schema.org/distribution> - A downloadable form of this dataset, potentially with different file formats or content. - <https://schema.org/sponsor> - A sponsor of the creative work or project. - <https://schema.org/temporalCoverage> - The temporal coverage of the data.

5.3 Dataverse-Specific Predicates

These predicates are used to describe datasets and related metadata within the Dataverse platform, which is a key component of the ODISSEI infrastructure:

Citation	and	Identification:	
https://dataverse.org/schema/citation/datasetContactAffiliation		- Affiliation of the dataset contact.	-
https://dataverse.org/schema/citation/datasetContactName		- Name of the dataset contact.	-
https://dataverse.org/schema/citation/producerLogoURL		- URL of the producer's logo.	-
https://dataverse.org/schema/citation/distributorLogoURL		- URL of the distributor's logo.	-
https://dataverse.org/schema/citation/distributor		- The distributor of the dataset.	-
https://dataverse.org/schema/citation/datasetContact		- Contact information for the dataset.	-
https://dataverse.org/schema/citation/dateOfCollection		- Date or dates when the data was collected.	-
https://dataverse.org/schema/citation/depositor		- The person or organization who deposited the dataset.	-
https://dataverse.org/schema/citation/distributionDate		- Date of distribution of the dataset.	-
https://dataverse.org/schema/citation/dsDescription		- Description of the dataset.	-
https://dataverse.org/schema/citation/keyword		- Keywords associated with the dataset.	-
https://dataverse.org/schema/citation/notesText		- General notes about the dataset.	-
https://dataverse.org/schema/citation/otherId		- Other identifiers for the dataset.	-
https://dataverse.org/schema/citation/producer		- The producer of the dataset.	-
https://dataverse.org/schema/citation/productionDate		- Date of production of the dataset.	-
https://dataverse.org/schema/citation/productionPlace		- Place of production of the dataset.	-
https://dataverse.org/schema/citation/series		- Series to which the dataset belongs.	-
https://dataverse.org/schema/citation/subtitle		- Subtitle of the dataset.	-
https://dataverse.org/schema/citation/topicClassification		- Topic classification for the dataset.	-
https://dataverse.org/schema/citation/accessToSources		- Information on how to access the sources of the data.	-
https://dataverse.org/schema/citation/originOfSources		- Origin of the data sources.	-

<https://dataverse.org/schema/citation/relatedMaterial> - Related materials to the dataset.

Core Dataverse Properties:

- <https://dataverse.org/schema/core#generatedBy> - Indicates the software or process that generated the data.
- <https://dataverse.org/schema/core#currentIngestedName> - The current ingested name of the file.
- <https://dataverse.org/schema/core#UNF> - Universal Numeric Fingerprint for the data file.
- <https://dataverse.org/schema/core#checksum> - Checksum of the data file.
- <https://dataverse.org/schema/core#datasetVersionId> - Version ID of the dataset.
- <https://dataverse.org/schema/core#fileTermsOfAccess> - Terms of access for the file.
- <https://dataverse.org/schema/core#filesize> - Size of the file in bytes.
- <https://dataverse.org/schema/core#restricted> - Indicates if the file is restricted.
- <https://dataverse.org/schema/core#rootDataFileId> - ID of the root data file.
- <https://dataverse.org/schema/core#storageIdentifier> - Storage identifier for the file.
- <https://dataverse.org/schema/core#termsOfUse> - Terms of use for the dataset.

5.4 ODISSEI Portal-Specific Predicates

These predicates are unique to the ODISSEI portal and describe specific aspects of social science data and its management:

Geospatial Information:

- <https://portal.odissei.nl/schema/geospatial#city> - City associated with the geospatial data.
- <https://portal.odissei.nl/schema/geospatial#geographicCoverage> - Geographic coverage of the dataset.
- <https://portal.odissei.nl/schema/geospatial#geographicBoundingBox> - Bounding box coordinates for geographic coverage.

DANS Metadata:

- <https://portal.odissei.nl/schema/dansMetadata#dansRightsHolder> - Rights holder according to DANS metadata.
- <https://portal.odissei.nl/schema/dansMetadata#dansAudience> - Target audience according to DANS metadata.
- <https://portal.odissei.nl/schema/dansMetadata#dansCessdaClassification> - CEESDA (Consortium of European Social Science Data Archives) classification.
- <https://portal.odissei.nl/schema/dansMetadata#dansCollection> - Collection information according to DANS metadata.
- <https://portal.odissei.nl/schema/dansMetadata#dansElsstClassification> - ELSST (European Language Social Science Thesaurus) classification.
- <https://portal.odissei.nl/schema/dansMetadata#dansMetadataLanguage> - Language

of the metadata. -
<https://portal.odissei.nl/schema/dansMetadata#dansPersonalDataPresent> -
 Indicates if personal data is present. -
<https://portal.odissei.nl/schema/dansMetadata#dansRelation> - Related resources
 according to DANS metadata. -
<https://portal.odissei.nl/schema/dansMetadata#dansTemporalCoverage> - Temporal
 coverage according to DANS metadata.

Enrichment and Linkage: -
<https://portal.odissei.nl/schema/enrichments#cbsLinkage> - Linkage to CBS
 (Statistics Netherlands) data. -
<https://portal.odissei.nl/schema/enrichments#cbsLinkageDetails> - Details of CBS
 linkage. - <https://portal.odissei.nl/schema/enrichments#term> - General term for
 enrichment.

Provenance and Workflow: - <https://portal.odissei.nl/schema/provenance#workflow>
 - Workflow associated with the data. - <https://www.w3.org/TR/prov-o/#wasDerivedFrom> -
 Indicates that an entity was derived from another. - <https://www.w3.org/TR/prov-o/#wasGeneratedBy> - Indicates the activity that generated an entity.

Question Information: -
<https://portal.odissei.nl/schema/questionInformation#questionSchemeName> - Name
 of the question scheme. -
<https://portal.odissei.nl/schema/questionInformation#question> - The question text.
 -
<https://portal.odissei.nl/schema/questionInformation#questionSchemeDescription>
 - Description of the question scheme.

Social Science Specifics: -
<https://portal.odissei.nl/schema/socialscience#samplingProcedure> - Description of
 the sampling procedure. -
<https://portal.odissei.nl/schema/socialscience#actionsToMinimizeLoss> - Actions
 taken to minimize data loss. -
<https://portal.odissei.nl/schema/socialscience#collectionMode> - Mode of data
 collection. -
<https://portal.odissei.nl/schema/socialscience#dataCollectionSituation> -
 Situation of data collection. -
<https://portal.odissei.nl/schema/socialscience#frequencyOfDataCollection> -
 Frequency of data collection. -
<https://portal.odissei.nl/schema/socialscience#responseRate> - Response rate of the

study. - <https://portal.odissei.nl/schema/socialscience#universe> - Universe of the study.
 study. - <https://portal.odissei.nl/schema/socialscience#weighting> - Weighting applied to the data.
<https://portal.odissei.nl/schema/socialscience#socialScienceNotes> - General notes on social science aspects.

Variable **Information:** -
<https://portal.odissei.nl/schema/variableInformation#odisseiVariable> - ODISSEI specific variable.

5.5 Other Relevant Predicates

- <http://rdf-vocabulary.ddialliance.org/discovery#kindOfData> - Kind of data according to DDI Alliance.
- <http://www.openarchives.org/ore/terms/describes> - Describes a resource.
- <http://www.w3.org/2002/07/owl#sameAs> - Indicates that two individuals are the same.
- <http://purl.org/dc/terms/bibliographicCitation> - Bibliographic citation for the resource.

6. Conclusion

This documentation provides a comprehensive overview of the ODISSEI Knowledge Graph RDF schema, highlighting its structure, key classes, and predicates. The schema's design, incorporating both standard vocabularies and domain-specific extensions, reflects the complexity and richness of social science data. Understanding this schema is fundamental for researchers and data scientists to effectively navigate, query, and utilize the valuable resources within the ODISSEI Knowledge Graph.

7. References

[1] ODISSEI Knowledge Graph SPARQL Endpoint:
<https://api.kg.odissei.nl/datasets/odissei/odissei-knowledge-graph-acceptance/services/odissei-knowledge-graph-acceptance-virtuoso/sparql>