-

Knowledge Graph Metadata Specification and Validation

Maryam MOHAMMADI [a,1] Chang SUN [a] Christopher BREWSTER [a,b]
Michel DUMONTIER [a]

[a] *Institute of Data Science, Department of Advanced Computing Sciences, Maastricht University, The Netherlands*

[b] *Data Science Group, TNO, Soesterberg, The Netherlands*

Abstract. Knowledge Graphs (KGs) are published across diverse repositories using heterogeneous metadata models that vary in their descriptive elements, level of detail, vocabularies, and support for semantic interoperability, in part due to the absence of a standardized specification. This lack of consistency in metadata practices hinders effective discovery and reuse of KGs. To address this gap, we present a KG Metadata Specification and Validation Framework designed to foster structured, FAIR-aligned KG descriptions. Developed through a community-driven process, the specification defines 33 metadata elements formalized using the Shapes Constraint Language (SHACL) to enable automated validation. We applied the framework to metadata from 1,573 datasets in the Linked Open Data (LOD) Cloud by mapping its 17 metadata fields to those in our specification. This mapping demonstrated the specification's compatibility with the largest Linked Data repository and validation process revealed common issues and areas for improvement in metadata quality. Furthermore, we demonstrated how metadata can be published in a way that enhances its visibility for end users and its discoverability by search engines, achieving full compliance with FAIR-Checker evaluations through the publication of specification-conformant metadata for the Food Health Claims KG. The framework provides a foundation for KG publishers, repository managers, and researchers, promoting uniform, high-quality metadata practices and advancing FAIR principles within the KG community.

Keywords. Knowledge graph, Metadata specification, SHACL validation, FAIR

1. Introduction

Knowledge Graphs (KGs) are data structures that represent real-world concepts or objects as nodes and their relationships as edges. They are useful for integrating and analyzing heterogeneous data [2]. KGs are becoming widely used across diverse domains such as education [16,17], healthcare [14], and artificial intelligence [4,18], and they are supporting systems like recommender engines [15], question answering [19], and information retrieval [20,21]. Their ability to represent rich, interconnected, and semantically diverse data makes KGs a powerful tool for data integration and analytics [23]. However, this structural complexity also introduces challenges in understanding, navigating, and

[1]Corresponding Author: Maryam Mohammadi, m.mohammadi@maastrichtuniversity.nl.

reusing their content effectively [24]. Addressing such complexity requires high-quality metadata that can guide users and systems in understanding the content of KG. However, current KG metadata remain inconsistent and often incomplete [12,13,22], which significantly hampers discovery, comparison, and reuse, and limits alignment with the FAIR principles [5].

Several efforts aim to describe datasets, including the Data Catalog Vocabulary (DCAT) [6], the Provenance Ontology (PROV) [7], the Dublin Core Metadata Initiative (DCMI) [8], the Vocabulary of Interlinked Datasets (VoID) [9], and Schema.org [10]. The Metadata for Ontology Description and publication (MOD) vocabulary [11] supports the reuse of ontologies by providing structured metadata tailored to ontology users and practitioners. The W3C Healthcare and Life Sciences (HCLS) Community Group has proposed a specification [1] that integrates multiple RDF vocabularies to better represent RDF datasets. While the HCLS model provides a strong foundation for RDF description, it does not explicitly address certain KG-specific features that are highly valuable in practice. Examples include metadata that provides a sample SPARQL query, giving users a concrete starting point for exploring a KG's structure and content, and metadata for a KG's REST API endpoint, which facilitates programmatic access and enables developers and applications to integrate and interact with the KG more easily.

Beyond these vocabularies, initiatives such as the Linked Open Data (LOD) Cloud[2] and the KGHub Knowledge Graph Registry[3] collect and organize metadata about existing KGs. While these vocabularies and registries address important aspects of dataset or ontology description, none provides a complete solution for the diverse requirements of KG metadata. This gap motivates the development of a more expressive and comprehensive approach to KG metadata specification.

In this work, we present a KG metadata specification and validation framework, informed by the HCLS specification and enriched with additional elements and vocabularies tailored to KG functionalities. Where applicable, vocabulary terms are mapped to their Schema.org equivalents to enable indexing and discovery by services such as Google Dataset Search, thereby enhancing dataset findability. The specification is developed through a community-driven process using a collaborative spreadsheet, and its elements and constraints are formalized into Shapes Constraint Language (SHACL) to support automated validation. We compare and map the proposed specification to the LOD Cloud metadata schema, applying SHACL validation to LOD Cloud datasets to identify common issues. Moreover, we demonstrate how the framework can be applied to create and publish KG metadata by showcasing the publication of a real-world KG from the Food Health Claims project [30]. The framework is useful for enhancing the quality of KG metadata, improving the discovery and aggregation of metadata, and ultimately fostering the reuse of KGs.

The remainder of this paper is organized as follows: Section 2 describes the method. Section 3 presents the specification and LOD Cloud metadata evaluation. Section 4 demonstrates the publication of Food Health Claims KG metadata in a FAIR-compliant manner. Section 5 compares the specification to existing models, interprets the results, and discusses limitations and future work. Finally, Section 6 summarizes the main contributions and findings.

[2] `https://lod-cloud.net/`
[3] `https://kghub.org/kg-registry/`

2. Method

The metadata specification was collaboratively developed during a three-day workshop[4] organized as part of the COST Action[5] on Distributed Knowledge Graphs (DKG). 13 Experts in KG development, metadata curation, and semantic technologies contributed to defining the essential elements needed to thoroughly describe KGs. The initial version was captured in a collaborative spreadsheet and later refined and formalized into SHACL [26] during the FedQuery Workshop[6] in 2024. The following describes the workflow that led to the final KG Metadata Specification and its associated validation shapes.

2.1. KG Metadata Specification

The aim was to define a set of metadata elements that provide a complete description of KGs while remaining aligned with existing schemas to maximize interoperability and adoption. The development process followed an iterative, consensus-based workflow:

- **Brainstorming KG-Specific Metadata Elements:** Experts collaboratively discussed the scope of information needed to comprehensively describe KGs for both human and machine use, focusing on aspects such as discoverability, interoperability, accessibility, and reusability in line with the FAIR principles. As part of this process, participants reviewed existing KG repositories and metadata models (e.g., HCLS, DCAT, VoID, MOD, Schema.org, and the LOD cloud) to identify common best practices and areas where KG-specific requirements were missing. This step revealed several gaps that were not well supported by existing models, including the need for metadata elements such as example SPARQL queries, REST API endpoints, visual representations of the KG schema, and fine-grained provenance roles.
- **Selecting Vocabularies and Data Types:** For each metadata element, we assigned an appropriate vocabulary IRI or data structure, reusing RDF terms from the reviewed schemas and vocabularies. Where a Schema.org equivalent was available, it was included alongside the primary RDF term to support broader adoption and enable indexing by services like Google Dataset Search.
- **Consensus and Minimum Set Definition:** Proposed elements were iteratively discussed and refined until consensus was reached. For each element, the purpose, data type, and cardinality were defined. In addition, we classified elements as mandatory or optional based on their relevance to core metadata functions such as identification, access, licensing, provenance, and interoperability, which are key aspects of the FAIR principles. The set of mandatory elements establishes the minimum requirements for a high-quality KG description.
- **Specification Illustration:** To test the utility and validity of the specification, we created metadata for Wikidata [25]. We chose Wikidata because it is widely used, openly accessible, and familiar to the community.

[4] `https://github.com/cost-dkg/metadata4dkg2022`
[5] `https://cost-dkg.eu/`
[6] `https://github.com/MaastrichtU-IDS/FedQuery2024`

2.2. KG Metadata Validation

To enable automated validation of KG metadata, each element in the proposed specification was formalized using SHACL [26]. These SHACL shapes define constraints such as cardinality and data types, establishing clear validation rules. To validate the SHACL shapes, we performed a manual test using a representative metadata instance based on Wikidata. We iteratively modified individual values and reran the validation to check whether constraint violations occurred as expected, and valid data passed successfully.

2.2.1. LOD Metadata Validation

To assess the specification's compatibility with established metadata practices, we map it to the Linked Open Data (LOD) Cloud metadata model[7][8]. The LOD Cloud is one of the largest publicly available collections of interlinked datasets on the web [13] and provides machine-readable metadata in JSON format, but its field names do not reuse shared RDF vocabulary terms. We therefore manually examined each field and its values across multiple KGs in the LOD Cloud to interpret their semantics and determine the most appropriate corresponding element in our specification, enabling a structured mapping between the two models.

Out of 17 fields in LOD Cloud metadata model, 13 had direct semantic matches: title, description, website, owner, contact_point, triples, license, namespace, DOI, image URL, keywords, domain, and SPARQL. The remaining four: example, links, full_download, and other_download required closer inspection and examination of their values. This mapping process was implemented in a Python script, which parsed the original JSON metadata and transformed it into a Turtle file containing RDF vocabularies. The resulting RDF metadata was then validated against our SHACL-based constraints using the PySHACL library [27], allowing us to assess its conformance with the proposed specification.

3. Results

3.1. KG Metadata Specification

The proposed KG Metadata Specification defines 33 core elements, reusing terms from 13 well-established vocabularies (Table 1). Each element is described by a set of attributes: whether it is mandatory, its cardinality, the associated RDF property, any mapping to Schema.org, the expected value type, its purpose, and an illustrative value drawn from Wikidata.

The complete specification is maintained in a collaborative spreadsheet. For clarity, Table 2 presents a streamlined version containing the most essential columns, while Figure 1 shows a fully detailed entry for the element dct:title. In this example, the cardinality constraint specifies that the title must appear at least once but may have multiple values (e.g., alternative or multilingual titles). The allowed data types are either a language-tagged string (rdf:LangString) or a simple string (xsd:string), and the purpose is to provide the formal name of the KG.

[7] `https://lod-cloud.net/add-dataset`
[8] `https://lod-cloud.net/lod-data.json`

Of the 33 elements, 14 are designated as mandatory (shown in bold in Table 2): rdf:type, dct:title, dct:identifier, dct:description, foaf:page, prov:qualifiedAttribution, dct:issued, void:vocabulary, dcat:distribution, dcat:version, dct:license, dcat:keyword, dct:language, and dct:accessRights. These elements were chosen for their essential role in supporting key aspects of the FAIR principles. For instance, dct:title and dct:description are fundamental to identifying the dataset, while dcat:distribution ensures users can access it. Provenance and licensing elements, such as prov:qualifiedAttribution, dct:license, and dct:issued, are critical for enabling responsible reuse and ensuring legal clarity. Other elements, such as void:vocabulary, dcat:keyword, and dct:language, support dataset contextualization, filtering, and interoperability, and are therefore mandatory for FAIR-aligned KG descriptions.

Table 1. Vocabularies reused in the metadata specification

Vocabulary Name	Prefix	Namespace URI
Dublin Core Terms (DCTerms)	dct:	`http://purl.org/dc/terms/`
Schema.org	schema:	`http://schema.org/`
Data Catalog Vocabulary (DCAT)	dcat:	`http://www.w3.org/ns/dcat#`
Provenance, Authoring and Versioning (PAV)	pav:	`http://purl.org/pav/`
Friend of a Friend (FOAF)	foaf:	`http://xmlns.com/foaf/0.1/`
Vocabulary of Interlinked Datasets (VoID)	void:	`http://rdfs.org/ns/void#`
Provenance Ontology (PROV)	prov:	`http://www.w3.org/ns/prov#`
Quantities, Units, Dimensions and Data Types (QUDT)	qudt:	`http://qudt.org/schema/qudt/`
Metadata for Ontology Description (MOD)	mod:	`https://w3id.org/mod#`
Asset Description Metadata Schema (ADMS)	adms:	`http://www.w3.org/ns/adms#`
Citation Typing Ontology (CiTO)	cito:	`http://purl.org/spar/cito`
RDF Schema (RDFS)	rdfs:	`http://www.w3.org/2000/01/rdf-schema#`
Resource Description Framework (RDF)	rdf:	`http://www.w3.org/1999/02/22-rdf-syntax-ns#`

Card.	Property	Datatype	Schema.org Term	Schema Datatype	Purpose	Wikidata Example
[1, ∞]	dct:title	rdf:LangString or xsd:string	schema:name	schema:Text	The name or formal title for the KG.	Wikidata@en

Figure 1. An example element from the specification spreadsheet.

Figure 2 illustrates the structural relationships among the components of the specification. At its core is the dcat:Dataset entity, which incorporates all 33 metadata elements proposed for describing KGs. Many of these elements are represented as IRIs or literals (e.g., strings or other datatypes), while others reference structured entities that allow more detailed descriptions. For example, void:sparqlEndpoint links to the dcat:DataService entity, which captures programmatic access points through properties such as dcat:endpointURL and dcat:endpointDescription. Likewise, the dcat:distribution element connects to the dcat:Distribution class, where additional details about downloadable files are specified, including their title, description, media type, download URL, and access URL.

A key design feature is the adoption of the flexible prov:qualifiedAttribution model. In contrast to simpler provenance fields that record only a dataset's creator, this model enables detailed attribution of multiple agents (e.g., owners, funders, authors, publishers) together with their specific roles, making use of the CI_RoleCode vocabulary[9]. This approach supports more precise credit assignment, increases transparency, and facilitates complex provenance tracking in KG publishing workflows.

Finally, void:Linkset provides a mechanism for explicitly linking datasets by specifying the target dataset along with the number of triples involved. A complete example of specification-conformant metadata for Wikidata is presented in Listing 2 in Appendix A.

Table 2.: Overview of all 33 metadata elements defined in the proposed KG metadata specification, with their descriptions, corresponding properties, and Schema.org vocabularies. Bold Fields are mandatory elements.

Field	**Purpose**	**Property**	**Data Type**	**Schema.org Term**	**Schema Data Type**
Identifier	The identifier for KG metadata.	dct:identifier	rdfs:Literal or IRI	schema:identifier	schema:Text or schema:URL
Type	The type of object in the description.	rdf:type	dcat:Dataset	rdf:type	schema:Dataset
Title	The name or formal title for the KG.	dct:title	rdf:LangString or xsd:string	schema:name	schema:Text
Alternative Title	Another name for the KG.	dct:alternative	rdf:langString or xsd:string	schema:alternateName	schema:Text
Acronym	An acronym used to identify the KG.	qudt:acronym	xsd:string	schema:termCode	schema:Text
Description	A human-readable description of the KG.	dct:description	rdf:LangString or xsd:string	schema:description	schema:Text
Homepage URL	The main web page for the KG.	foaf:page	IRI	schema:url	schema:URL
Other Pages	Additional relevant pages for the KG.	rdfs:seeAlso	IRI	schema:relatedLink	schema:URL
Roles [e.g., Creator, Contact point,Funder, Publisher,]	relating a resource to another resource with a specified role.	prov:qualifiedAttribution → prov:agent —-→foaf:name —-→foaf:mbox → dcat:hadRole	prov:Atribution → prov:Agent [IRI] → CI_RoleCode	→ schema:creator → schema:contactPoint → schema:funding →schema:publisher	schema:Organization or schema:Person → schema:name → schema:email →schema:publisher
Created Date	The date when the KG was created.	pav:createdOn or dct:created	xsd:dateTimeStamp	schema:dateCreated	schema:Date or schema:DateTime
Modified Date	The last modification date of the KG.	dct:modified	xsd:dateTimeStamp	schema:dateModified	schema:Date or schema:DateTime
Published Date	The date when the KG was published.	dct:issued	xsd:dateTimeStamp	schema:datePublished	schema:Date or schema:DateTime
Vocabularies Used	The vocabularies used in the KG.	void:vocabulary	IRI	-	-

[9] `https://standards.iso.org/iso/19115/resources/Codelists/gml/CI_RoleCode.xml`

Field	**Purpose**	**Property**	**Data Type**	**Schema.org Term**	**Schema Data Type**
Primary Reference Document	A document or data paper describing the KG.	cito:citesAsAuthority	IRI	schema:subjectOf	schema:CreativeWork or schema:Event
Meta Graph (Picture)	Illustration of the KG connections.	foaf:depiction	IRI	schema:image	schema:ImageObject or schema:URL
Statistics	Computational representation of the KG.	VOID vocabularies, e.g. void:triples	various data types, e.g. rdfs:Literal for void:triples	-	-
KG schema	A formal specification to validate the KG.	dct:conformsTo	IRI	-	-
Distributions	A specific representation of the KG.	dcat:distribution	dcat:Distribution → dct:title → dct:description → dcat:mediaType → dcat:accessURL → dcat:byteSize → dct:format → dcat:hasPolicy → dct:license → dct:issued → dct:rights → dct:modified	schema:distribution	schema:DataDownload → schema:name → schema:description → schema:contentUrl → schema:contentSize → schema:license → schema:datePublished → schema:usageInfo → schema:dateModified
REST API	REST API for the KG.	inverse (dcat:servesDataset)	dcat:DataService → dcat:endpointURL → dct:identifier → dct:title → adms:status → dcat:servesDataset	schema:mainEntityOfPage	schema:WebAPI → schema:endpointURL → schema:identifier → schema:name → schema:dataset
SPARQL Endpoint	SPARQL endpoint for querying the KG.	void:sparqlEndpoint	IRI	schema:contentUrl	schema:URL
Example Queries	Example queries against the KG.	mod:sampleQueries	Literal or IRI	-	-
Version	The KG version.	dcat:version	Literal	schema:version	schema:Number or schema:Text
License	License under which the KG is released.	dct:license	IRI	schema:license	schema:CreativeWork or schema:URL
Keywords	Keywords related to the KG.	dcat:keyword	xsd:string	schema:keywords	schema:DefinedTerm or schema:Text
Category	A main category of the resource.	dcat:theme	IRI	schema:category	schema:CategoryCode
References	Any document related to the KG.	dct:references	IRI	schema:publication, schema:url,schema:about	schema:PublicationEvent
Language	Languages represented in the KG.	dct:language	rdf:langString IETF language tag (bcp-47)	schema:inLanguage	schema:Language
IRI Template	IRI Templates following RFC 6570.	void:uriRegexPattern or hydra:template	xsd:string	-	-
Linked Resources	Linkable resources to the KG.	inverse(void:target)	void:LinkSet → void:target → void:triples	-	-

Field	Purpose	Property	Data Type	Schema.org Term	Schema Data Type
Example Resource	An example instance from the KG.	void:exampleResource	dcat:Resource → dct:title → dct:description → adms:status → void:accessURL	-	-
Access Statement	Access restrictions on the KG.	dct:accessRights	IRI, dct:RightsStatement	-	-
Source	The origin or source of data for the KG.	prov:hadPrimarySource	IRI	-	-
name space	To state that all entity URIs in a dataset start with a given string.	void:uriSpace	literal	-	-

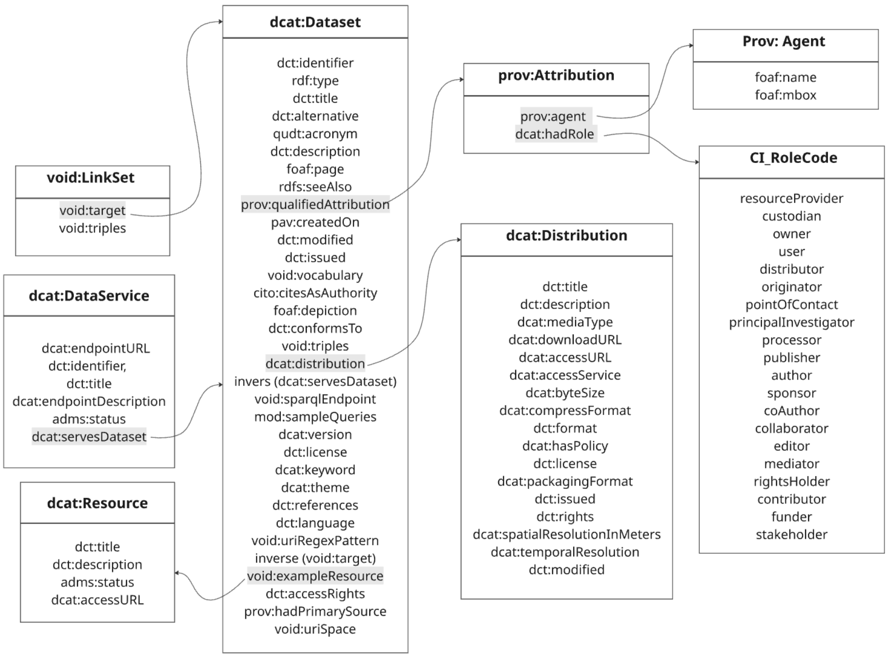

Figure 2. Structural relationships between components of specification .

3.2. Metadata Validation

To enable automated validation of metadata elements and their relationships, all 33 elements in the specification were formalized using SHACL. For each shape, SHACL constraints define the property path (sh:path), expected datatype (sh:datatype), RDF node kind (sh:nodeKind), and cardinality (sh:minCount, sh:maxCount), as well a human-readable description to support implementers (:example). Listing 1 illustrates an example using the dct:title property, which must appear at least once (sh:minCount 1) and must have a value that is either a plain string (xsd:string) or a language-tagged string (rdf:langString).

As shown in Figure 2, several metadata elements (e.g., dcat:distribution, void:sparqlEndpoint, prov:qualifiedAttribution) reference structured entities. These were modeled as separate SHACL shapes to capture their internal properties and constraints. The complete SHACL implementation[10] defines four shapes: DCATDatasetShape, which covers the full set of 33 metadata elements describing a KG; AgentShape, which captures roles and responsibilities via prov:qualifiedAttribution; DCATDataService, which describes programmatic access points such as SPARQL endpoints; and LinksetShape, which specifies dataset-to-dataset links. Each shape constrains the properties relevant to its corresponding entity. For example, AgentShape validates attributes such as foaf:name and foaf:mbox; DCATDataService includes dcat:endpointURL and dcat:endpointDescription; and LinksetShape requires void:target and void:triples.

```
:DCATDatasetShape
    a sh:NodeShape ;
    sh:targetClass dcat:Dataset ;
    sh:property [
        sh:path dct:title ;
        sh:minCount 1 ;
        sh:or (
           [ sh:datatype rdf:langString ]
           [ sh:datatype xsd:string ]
              ) ;
        sh:description "To provide a name for the KG." ;
        sh:example "Wikidata"^^xsd:string ;
        :example "Wikidata"@en ;
                ] ;
```

Listing 1: SHACL Shape Example

3.2.1. LOD metadata validation

To demonstrate the utility of the framework, we validate LOD cloud metadata against SHACL-based metadata specifications. The 2025 LOD cloud contains 1,573 datasets and includes 17 metadata fields in its catalog. As described in Section 2.2.1, we map elements of the LOD metadata schema to the vocabularies defined in our specification and generate a Turtle file compatible with the validation framework.

Table 3 summarises the mapping between LOD metadata fields and the proposed specification. For each element, it also provides the total number of instances, the number of datasets in which it appears at least once, and the corresponding number of validation violations. The validation identified a total of 22,349 violations. The majority of these, 16,507 cases, were due to missing mandatory values ("Less than 1 values"). For example, the dct:license element was present in 835 datasets but missing from 738, resulting in 738 validation errors. Other frequently missing mandatory elements included foaf:page, dcat:keyword, foaf:name and foaf:mbox (for owner and contact point roles), as well as distribution-related fields such as dcat:downloadURL, dcat:accessURL, dcat:mediaType, and descriptive fields such as dct:title, dct:description, dcat:version, dct:language, dcat:distribution, dct:issued, and void:vocabulary.

[10] https://github.com/MaastrichtU-IDS/kg-metadata/blob/main/shacl/full-hcls_shacl.ttl

The remaining 5,842 violations were caused by structural mismatches (“Value does not conform to Shape”). These occurred in composite elements where nested structures did not match the specified constraints. The most frequent sources of violations were Distribution, Owner, and Contact Point entities, which lacked required internal properties. All conformance errors were linked to missing mandatory sub-elements, and in such cases SHACL reported violations both at the sub-element and at the parent shape level. No datatype-related violations were observed, because during the conversion from JSON to Turtle, malformed URIs, incorrect data types, and other syntax errors were corrected, while missing values were not imputed.

Table 3.: Mapping of LOD metadata elements to specification fields. #Instances indicates the total number of occurrences of each metadata element across the LOD cloud. #DS refers to the number of datasets in which the element appears at least once. #Violations represents the number of validation errors associated with each corresponding metadata element.

LOD Field	**#Instances [#DS]**	**Specification Field**	**Vocabulary**	**#violations**
Title	1573 [1573]	Title	dct:title	0
Description	1573 [1573]	Description	dct:description	0
Website	1255 [1255]	Homepage URL	foaf:page	318
Triples	1568 [1568]	Statistics	void:triples	0
License	835 [835]	License	dct:license	738
Namespace	705 [705]	Namespace of URIs	void:uriSpace	0
DOI	17 [17]	Identifier	dct:identifier	0
Image URL	67 [67]	Meta Graph (Picture)	foaf:depiction	0
Keywords	15743 [1493]	keywords	dcat:keyword	80
Domain	1244 [1244]	category	dcat:theme	0
SPARQL	668 [657]	SPARQL Endpoint	void:sparqlEndpoint	0
→ Media_Type	511	*→ Media Type*	dcat:mediaType	0
→ Title	451	*→ Title*	dct:title	0
→ Access_URL	668	*→ Access URL*	dcat:endpointURL	0
→ Status	668	*→ Status*	adms:status	0
→ Id	139	*→ Identifier*	dct:identifier	0
→ Description	480	*→ Description*	dcat:endpointDescription	0
Example	1173 [828]	Example Resource	void:exampleResource	0
→ Title	607	*→ Title*	dct:title	0
→ Media_Type	1136	*→ Media Type*	dcat:mediaType	0
→ Access_URL	1173	*→ Access URL*	dcat:accessURL	0
→ Status	1173	*→ Status*	adms:status	0
→ Description	956	*→ Description*	dct:description	0

Continued on next page

LOD Field	#Instances [#DS]	Specification Field	Vocabulary	#violations
Owner	298 [298]	owner	prov:qualifiedAttribution	1553
→ Name	20	→ *Name*	foaf:name	540
→ Email	290	→ *Email*	foaf:mbox	0
Contact_point	2276 [1573]	Contact Point	prov:qualifiedAttribution	623
→ Name	1323	→ *Name*	foaf:name	253
→ Email	953	→ *Email*	foaf:mbox	993
Links	17882 [1295]	Linked Resources	void:LinkSet	0
→ target	17882	→ *target*	inverse(void:target)	0
→ value	17880	→ *triple*	inverse(void:triples)	0
Full_Download & other_download	512/3544 [282/1090]	Distributions	dcat:distribution	4403
→ Media_Type	511/3464	→ *Media Type*	dcat:mediaType	1
→ Description	415/2776	→ *Description*	dct:description	14
→ Title	449/2179	→ *Title*	dct:title	8
→ Id	267/62	→ *Identifier*	dct:identifier	0
→ status	512/3544	→ *Status*	adms:status	0
→ Download_URL	511	→ *Download URL*	dcat:downloadURL	3545
→ Access_URL	3544	→ *Access URL*	dcat:accessURL	522

4. Demonstration

4.1. Publication of KG Metadata

Creating and validating high-quality metadata is an essential step toward improving the FAIRness of a KG. To further enhance FAIRness, metadata should be made accessible through channels that support both human-readable presentation and machine-readable indexing. In this section, we apply the proposed Metadata Specification and Validation Framework to the Food Health Claims KG[11] [30] and describe how the resulting metadata was published to enhance its visibility for end users and its discoverability by search engines. This KG, developed by one of the authors, captures structured representations of 260 authorized health claims, based on scientific assessments of food ingredients and their health effects.

The metadata for the Food Health Claims KG was created and validated using the proposed specification. We primarily used terms from the schema.org vocabulary to maximize visibility through web-based indexing services, and supplemented these with additional RDF vocabularies to fully represent all 33 metadata elements defined in our framework. The resulting metadata was expressed in JSON-LD format and embedded in an HTML page that also provides a human-readable overview of the KG. This page was deployed using GitHub Pages under a publicly accessible URL[12]. As recommended by the

[11] `https://github.com/MaastrichtU-IDS/food-claims-kg/tree/gh-pages`
[12] `https://maastrichtu-ids.github.io/food-claims-kg/`

FAIR principles, and specifically sub-principle F1 [5], the metadata includes a globally unique and persistent identifier. To implement this, a dedicated w3id.org namespace[13] was registered. This persistent URI redirects to the metadata page, allowing the value used in the schema:identifier field to remain stable over time, even if the hosting location changes. We then validated the generated metadata using SHACL[14] to ensure conformance with the specification. The complete script containing specification-conforming metadata for Food Health Claims KG is provided in the listing 3 in Appendix B.

To verify whether search engines can correctly interpret the embedded structured data, the metadata page was tested using Google's Rich Results Test[15]. After successful validation, the site was registered in Google Search Console[16] to confirm site ownership and request indexing. Within a few days, the Food Health Claims KG metadata appeared among the top results in Google Dataset Search[17], significantly improving its discoverability. The scripts and deployment instructions are available in the project's GitHub repository.

4.2. KG Metadata FAIRness Assessment

In addition to enhancing web-based discoverability, the published metadata was evaluated using the FAIR-Checker service[18] [28,29], which automatically assesses metadata against FAIR principles [5]: Findable (F), Accessible (A), Interoperable (I), and Reusable (R). Each principle is divided into numbered sub-principles, such as F1 or F2, that define specific requirements including the use of persistent identifiers or the inclusion of structured metadata. FAIR-Checker further divides these into individual machine-checkable tests, identified by adding a letter suffix (for example, F1A or F1B), where each letter corresponds to a distinct criterion under the same sub-principle.

The Food Health Claims KG metadata achieved a score of 100% (12/12), with all tested criteria satisfied. The criteria covered all four FAIR principles: for findability, unique identifiers (F1A), persistent identifiers (F1B), structured metadata (F2A), and use of shared vocabularies (F2B); for accessibility, open resolution protocols (A1.1) and authorization or access statements (A1.2); for interoperability, machine-readable formats (I1), shared ontologies (I2), and external links (I3); and for reusability, a licence (R1.1), provenance information (R1.2), and compliance with community standards (R1.3) were provided. These tool-generated metrics confirm that the proposed specification and publication approach satisfies the FAIR principles.

5. Discussion

This study introduced a metadata specification for KGs, consisting of 33 elements formalized using SHACL to enable automated validation. The design process was community-driven, involving experts in KG development to establish best practices in-

[13] `https://w3id.org/foodhkg`
[14] https://github.com/MaastrichtU-IDS/food-claims-kg/blob/gh-pages/schema_shacl.ttl
[15] `https://search.google.com/test/rich-results`
[16] `https://search.google.com/search-console`
[17] `https://datasetsearch.research.google.com/`
[18] `https://fair-checker.france-bioinformatique.fr/`

formed by the FAIR principles [5] and the analysis of existing metadata practices and Schemas. The demonstration with the Food Health Claims KG illustrated that the specification can be readily implemented by KG publishers and that it produces metadata which achieves full compliance with FAIR principles according to FAIR-Checker evaluations, as shown in Section 4.2.

The FAIR-Checker evaluation showed that the metadata produced with our specification satisfies all machine-checkable FAIR criteria. However, FAIR-Checker currently applies a predefined set of generic tests and does not consider KG-specific validation rules. Our SHACL shapes help to address this gap by enabling more detailed checks on the structure and content of metadata. In this way, FAIR-Checker and SHACL are complementary: the former verifies compliance with high-level FAIR principles, while the latter ensures conformance with a community-agreed schema. Looking ahead, integrating SHACL-based validation into FAIR assessment services such as FAIR-Checker could offer a more comprehensive approach that combines general FAIR criteria with domain-specific requirements.

Compared to existing metadata schemas, our specification offers broader descriptive coverage while maintaining compatibility with established models. Validation against LOD Cloud metadata highlights three key aspects: (i) the specification fully accommodates all 17 metadata fields defined in the LOD Cloud schema, (ii) the associated SHACL schema is effective in detecting omissions and errors in LOD descriptions, and (iii) the specification introduces 16 additional metadata elements that extend beyond LOD Cloud's coverage. These additions include fields such as Type, Alternative Title, Acronym, Other Pages, a more expressive attribution model for roles (e.g., creator, owner, publisher), and temporal metadata such as Created, Modified, and Published dates. Further additions include Reference Article, KG Schema, REST API, Version, Language, IRI Template, Access Statement, and Source Datasets.

Compared to the HCLS profile, which provides a solid foundation for RDF dataset metadata and supports validation through ShEx, our specification extends coverage by introducing several practical elements that HCLS does not include. These include support for example SPARQL queries, REST API endpoints, visual representations of the KG, dataset acronyms, and fine-grained provenance roles. By incorporating these features, the specification enables richer metadata descriptions, improves programmatic access, and enhances transparency in attribution, while maintaining interoperability with existing standards. In practical terms, adoption of the proposed specification offers a standards-based template for both new KGs and the enhancement of existing ones. Its compatibility with established schemas lowers adoption barriers while its extended coverage enables more effective resource discovery and reuse.

The proposed specification provides a general and comprehensive set of metadata elements; however, certain use cases may still fall outside its scope. For instance, more technical information related to the methods used to create a KG, such as specific tools, workflows, or pipelines, is not explicitly captured. In addition, the specification does not currently include an element to indicate the underlying data model of a KG (for example, property graph or RDF). This reflects a decision to prioritize the reuse of existing vocabularies rather than creating new ones. Another limitation is the continued reliance on manual metadata creation and publication, which may hinder the adaptability, accuracy, and consistency of the specification.

Future work will address these challenges by developing an automated pipeline for metadata creation, validation, and publication, enabling decentralized KG metadata that can be harvested into KG registries. In addition, we plan to investigate the use of large language models (LLMs) and retrieval-augmented generation (RAG) for automated metadata generation. Future work will also focus on broader community adoption and formal standardization of the specification. In particular, seeking endorsement through the W3C community and working groups would strengthen its status as a recognized standard for KG metadata. This endorsement would promote interoperability, encourage widespread adoption, and facilitate integration with existing web and linked data best practices.

6. Conclusion

In this paper, we introduced a Metadata Specification and Validation Framework designed to establish a robust foundation for standardized KG metadata practices and publication. Developed through a community-driven process, the framework defines 33 metadata elements and encodes them, along with their constraints, using SHACL shapes to enable automated metadata validation.

To evaluate its applicability, we mapped the LOD Cloud metadata to the proposed specification, validated it with SHACL, and analyzed the resulting violations. This assessment revealed gaps in existing LOD Cloud metadata, underscoring the need for richer and more consistent descriptions of KGs. As a case study, we applied the framework to the Food Health Claims KG, demonstrating how publishers can implement the specification to generate metadata that is indexable by Google and thereby improve discoverability. Evaluation with FAIR-Checker confirmed that metadata produced in this way achieves full compliance with the FAIR principles.

This work advances the state of practice in KG metadata by offering a specification that is interoperable with established models, extends their descriptive coverage, and provides verifiable FAIR compliance. As future work, we plan to develop an intuitive user interface for the specification, investigate the use of LLMs and RAG for automated metadata generation, and seek community and W3C endorsement to formalize the specification as a recognized standard. Through these efforts, we aim to foster the broader adoption of high-quality, FAIR-aligned metadata across the KG ecosystem.

Data and Code Availability

All resources used in this study, including the KG metadata spreadsheet, SHACL scripts, LOD mapping and evaluation scripts, and metadata publication instructions, are openly available at [19].

Acknowledgment

This project has received funding from COST Action. Funding for M. Mohammadi is provided by the European Union's Horizon 2020 research and innovation program un-

[19] `https://github.com/MaastrichtU-IDS/kg-metadata`

der the Marie Skłodowska-Curie grant agreement No. 860801, as well as support from COST Action. We gratefully thank the participants of the workshop who contributed to the development of the metadata specification through their feedback and discussion: Manuel Paneque, Matthijs Sloep, Ilan Kernerman, Jinzhou Yang, Maxime Lefrançois, Katja Hose, Flavio De Paoli, Remzi Celebi, Erkan Yasar, and María del Mar Roldán. In particular, we are thankful to Maxime Lefrançois for his valuable and insightful comments, which significantly improved the quality and clarity of this work.

References

[1] Dumontier M, Gray AJG, Marshall MS, Alexiev V, Ansell P, Bader G, Baran J, Bolleman JT, Callahan A, Cruz-Toledo J, et al. The health care and life sciences community profile for dataset descriptions. PeerJ. 2016;4:e2331.

[2] Hogan A, Blomqvist E, Cochez M, d'Amato C, De Melo G, Gutierrez C, Kirrane S, Labra Gayo JE, Navigli R, Neumaier S, et al. Knowledge graphs. ACM Computing Surveys (Csur). 2021;54(4):1–37.

[3] Ji S, Pan S, Cambria E, Marttinen P, Yu PS. A survey on knowledge graphs: Representation, acquisition, and applications. IEEE Transactions on Neural Networks and Learning Systems. 2021;33(2):494–514.

[4] Peng C, Xia F, Naseriparsa M, Osborne F. Knowledge graphs: Opportunities and challenges. Artificial Intelligence Review. 2023;56(11):13071–13102.

[5] Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, Blomberg N, Boiten JW, da Silva Santos LB, Bourne PE, et al. The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data. 2016;3(1):1–9.

[6] World Wide Web Consortium (W3C). Data Catalog Vocabulary (DCAT). 2014. Available from: `https://www.w3.org/TR/vocab-dcat/`. Accessed 2024-11-30.

[7] World Wide Web Consortium (W3C). Provenance Ontology (PROV). 2013. Available from: `https://www.w3.org/TR/prov-o/`. Accessed 2024-11-30.

[8] Dublin Core Metadata Initiative. Dublin Core Metadata Terms. 2012. Available from: `https://www.dublincore.org/specifications/dublin-core/dcmi-terms/`. Accessed 2024-11-30.

[9] World Wide Web Consortium (W3C). Vocabulary of Interlinked Datasets (VoID). 2011. Available from: `https://www.w3.org/TR/void/`. Accessed 2024-11-30.

[10] Schema.org. Schema.org. 2024. Available from: `https://schema.org/`. Accessed 2024-11-30.

[11] Biswanath Dutta, Durgesh Nandini, and Gautam Kishore Shahi, MOD: metadata for ontology description and publication, *Proceedings of the International Conference on Dublin Core and Metadata Applications*, Dublin Core Metadata Initiative, 2015.

[12] Mohammadi M. (Semi-) Automatic Construction of Knowledge Graph Metadata. In: The Semantic Web: ESWC 2022 Satellite Events: Hersonissos, Crete, Greece, May 29–June 2, 2022, Proceedings. Springer; 2022. p. 171–178.

[13] Assaf A, Troncy R, Senart A. What's up LOD cloud? Observing the state of linked open data cloud metadata. In: The Semantic Web: ESWC 2015 Satellite Events: ESWC 2015 Satellite Events, Portorož, Slovenia, May 31–June 4, 2015, Revised Selected Papers 12. Springer; 2015. p. 247–254.

[14] Hejie Cui, Jiaying Lu, Shiyu Wang, Ran Xu, Wenjing Ma, Shaojun Yu, Yue Yu, Xuan Kan, Chen Ling, Joyce Ho, *et al.*, A survey on knowledge graphs for healthcare: Resources, applications, and promises, *arXiv preprint arXiv:2306.04802*, 2023.

[15] Qingyu Guo, Fuzhen Zhuang, Chuan Qin, Hengshu Zhu, Xing Xie, Hui Xiong, and Qing He, A survey on knowledge graph-based recommender systems, *IEEE Transactions on Knowledge and Data Engineering*, vol. 34, no. 8, pp. 3549–3568, 2020.

[16] Yousra Fettach, Mounir Ghogho, and Boualem Benatallah, Knowledge graphs in education and employability: A survey on applications and techniques, *IEEE Access*, vol. 10, pp. 80174–80183, 2022.

[17] Bilal Abu-Salih and Salihah Alotaibi, A systematic literature review of knowledge graph construction and application in education, *Heliyon*, vol. 10, no. 3, 2024.

[18] Freddy Lecue, On the role of knowledge graphs in explainable AI, *Semantic Web*, vol. 11, no. 1, pp. 41–51, 2020.

[19] Apoorv Saxena, Soumen Chakrabarti, and Partha Talukdar, Question answering over temporal knowledge graphs, *arXiv preprint arXiv:2106.01515*, 2021.

[20] Laura Dietz, Alexander Kotov, and Edgar Meij, "Utilizing knowledge graphs for text-centric information retrieval," in *Proceedings of the 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*, pp. 1387–1390, 2018.

[21] Ridho Reinanda, Edgar Meij, Maarten de Rijke, *et al.*, "Knowledge graphs: An information retrieval perspective," *Foundations and Trends® in Information Retrieval*, vol. 14, no. 4, pp. 289–444, 2020.

[22] Jeremy Debattista, Christoph Lange, Sören Auer, and Dominic Cortis, Evaluating the quality of the LOD cloud: An empirical investigation, *Semantic Web*, vol. 9, no. 6, pp. 859–901, 2018.

[23] Saher M. Mohamed, Saleh Farah, Abdelrahman Mahmoud Lotfy, Kareem Ayman Rizk, Abdelrahman Yasser Saeed, Shahenda Hatem Mohamed, Ghada Khoriba, and Tamer Arafa, Knowledge Graphs: The Future of Data Integration and Insightful Discovery, in *Advanced Research Trends in Sustainable Solutions, Data Analytics, and Security*, pp. 99–146, IGI Global Scientific Publishing, 2025.

[24] Nandish Jayaram, Arijit Khan, Chengkai Li, Xifeng Yan, and Ramez Elmasri, Querying knowledge graphs by example entity tuples, *IEEE Transactions on Knowledge and Data Engineering*, vol. 27, no. 10, pp. 2797–2811, 2015.

[25] D. Vrandečić and M. Krötzsch, Wikidata: A free collaborative knowledgebase, *Communications of the ACM*, 57(10):78–85, 2014, `https://doi.org/10.1145/2629489`

[26] W3C. Shapes Constraint Language (SHACL). W3C Recommendation, 20 July 2017. `https://www.w3.org/TR/shacl/`

[27] Ashley Sommer. PySHACL: Python validator for SHACL. `https://github.com/RDFLib/pySHACL`, 2019.

[28] Alban Gaignard, Thomas Rosnet, Frédéric De Lamotte, Vincent Lefort, and Marie-Dominique Devignes, FAIR-Checker: supporting digital resource findability and reuse with Knowledge Graphs and Semantic Web standards, *Journal of Biomedical Semantics*, vol. 14, no. 1, p. 7, Springer, 2023.

[29] Thomas Rosnet, Vincent Lefort, Marie-Dominique Devignes, and Alban Gaignard, FAIR-Checker, a web tool to support the findability and reusability of digital life science resources, In: *Journées Ouvertes en Biologie, Informatique et Mathématiques (JOBIM 2021)*, Paris, France, 2021.

[30] Celebi R, van Lier I, de Boer A, Dumontier M. Construction of a knowledge graph for food health claims. In: CEUR Workshop Proceedings. Vol. 3890. Aachen: CEUR-WS; 2024. p. 165–166.

Appendix A. Wikidata Metadata

The following listings provide complete metadata examples, one in RDF/Turtle and one in HTML/Schema.org, demonstrating compliance with the proposed specification

```
@prefix : <http://example.org/> .
@prefix dct: <http://purl.org/dc/terms/> .
@prefix schema: <http://schema.org/> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix wd: <http://www.wikidata.org/entity/> .
@prefix dcat: <http://www.w3.org/ns/dcat#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix pav: <http://purl.org/pav/> .
@prefix void: <http://rdfs.org/ns/void#> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix prov: <http://www.w3.org/ns/prov#> .
@prefix qudt: <http://qudt.org/schema/qudt/>.
@prefix mod: <https://w3id.org/mod#>.
@prefix adms: <http://www.w3.org/ns/adms#> .
@prefix cito: <http://purl.org/spar/cito> .
:wikidata
    a dcat:Dataset;
    dct:identifier "https://www.wikidata.org/entity/Q2013"^^xsd:
        string ;
    dct:title "Wikidata"@en ;
    dct:alternative ""@en ;
    qudt:acronym ""^^xsd:string ;
    dct:description """ Wikidata is a free and open knowledge base
        that can be read and edited by ..."""@en ;
    foaf:page <https://www.wikidata.org> ;
    rdfs:seeAlso <https://www.wikidata.org/wiki/Wikidata:Introduction
        > ;
    pav:createdOn "2012-10-29T15:30:00Z"^^xsd:dateTimeStamp ;
    dct:modified "2024-11-25T15:30:00Z"^^xsd:dateTimeStamp ;
    dct:issued "2012-10-29T15:30:00Z"^^xsd:dateTimeStamp ;
    void:vocabulary <https://github.com/mkroetzsch/wda/blob/master/
        wikidata-ontology.owl> ;
    cito:citesAsAuthority <https://dl.acm.org/doi/10.1145/2629489> ;
    foaf:depiction <https://www.researchgate.net/figure/A-simplified-
        class-level-diagram-of-the-Wikidata-knowledge-graph-for-
        biomedical-entities_fig1_339985351> ;
    void:triples "17841499814"^^xsd:integer ;
    dcat:distribution [
                        dct:title "latest dump of Wikidata";
                        dct:description "latest dump of Wikidata in
                            turtle format";
                        dcat:mediaType "application/octet-stream";
                        dcat:downloadURL <https://www.wikidata.org/
                            wiki/Wikidata:Database_download/nl>;
                            dcat:accessURL <> ;];
    void:sparqlEndpoint[
                        dcat:endpointURL <https://query.wikidata.org/
                            sparql> ;
                        dct:identifier "Wikidata SPARQL endpoint"^^
                            rdfs:Literal ;
                        dct:title "Wikidata Query Service"^^rdfs:
                            Literal ;
```

```
                         dcat:endpointDescription  """The Wikidata
                             Query Service generally contains the full
                             data of Wikidata,...""" ;
                        adms:status  <http://purl.org/adms/status/
                            Completed> ;];
    mod:sampleQueries <https://www.wikidata.org/wiki/Wikidata:
        SPARQL_query_service/queries/examples> ;
    dct:version "1.0"^^rdfs:Literal ;
    dct:license <https://www.wikidata.org/wiki/Q6938433> ;
    dcat:keyword "knowledge graph, wikidata, ontology"^^xsd:string ;
    dcat:theme <open_knowledge_graph>
    dct:references <https://www.wikidata.org/wiki/Wikidata:
        Introduction> ;
    dct:language "multiple languages" ;
    void:uriRegexPattern "https://www.wikidata.org/wiki/{id}"^^xsd:
        string ;
    void:exampleResource <https://www.wikidata.org/wiki/Q2013> ;
    dct:accessRights <https://www.wikidata.org/wiki/Q232932> ;
    prov:hadPrimarySource <https://huggingface.co/datasets/wikimedia/
        wikipedia> ;
    prov:qualifiedAttribution [
        prov:agent  :OwnerAgent ;
        dcat:hadRole :owner ;] ;
    prov:qualifiedAttribution [
        prov:agent :contactAgent ;
        dcat:hadRole :contact_point ;]
     .

:eurovoc a void:Linkset ;
    void:target :wikidata ;
    void:triples "3068"^^xsd:integer ;
 .

:OwnerAgent a prov:Agent ;
     foaf:name "Lucas_Werkmeister"^^xsd:string ;
     foaf:mbox <mailto:lucas.werkmeister@wikimedia.de> ;
 .

:contactAgent a prov:Agent ;
     foaf:name "Lucas_Werkmeister"^^xsd:string ;
     foaf:mbox <mailto:wikid@a@wikimedia.de> ;
 .

<https://www.wikidata.org/wiki/Q42722531> a dcat:DataService;
                              dcat:servesDataset   :wikidata.
```

Listing 2: Wikidata metadata compliant to the proposed specification

Appendix B. Full HTML Script

```
<!DOCTYPE html>
<html lang="en">
<head>
  <meta charset="UTF-8" />
```

```
  <!-- Google Search Console Verification -->
  <meta name="google-site-verification" content="googledf60f41fc68f0bea" />
  <title>Food Health Claims Knowledge Graph</title>
  <meta name="viewport" content="width=device-width, initial-scale=1" />

  <script type="application/ld+json">
{
  "@context": {
    "schema": "http://schema.org/",
    "dcat": "http://www.w3.org/ns/dcat#",
    "dct": "http://purl.org/dc/terms/",
    "foaf": "http://xmlns.com/foaf/0.1/",
    "prov": "http://www.w3.org/ns/prov#",
    "void": "http://rdfs.org/ns/void#",
    "mod": "https://w3id.org/mod#",
    "adms": "http://www.w3.org/ns/adms#",
    "xsd": "http://www.w3.org/2001/XMLSchema#",
    "rdf": "http://www.w3.org/1999/02/22-rdf-syntax-ns#",
    "rdfs": "http://www.w3.org/2000/01/rdf-schema#"
  },
  "@graph": [
    {
      "@id": "http://www.w3id.org/foodhkg",
      "@type": "schema:Dataset",
      "schema:identifier": "https://w3id.org/foodhkg",
      "schema:name": "Food Health Claims Knowledge Graph",
      "schema:alternateName": "Food Claims Knowledge Graph",
      "schema:termCode": "FHCKG",
      "schema:description": "FoodKG is a structured knowledge graph curated from 260 authorised EU food health claims, covering ingredients, effects, target populations, and scientific evidence, designed to support personalised nutrition and health applications.",
      "schema:url": { "@id": "https://github.com/MaastrichtU-IDS/food-claims-kg" },
      "schema:relatedLink": { "@id": "https://docs.google.com/spreadsheets/d/1RWZ6A1GB8m7PO5kjsbbbeI4ETLwvKLOvkrzOpl8zAM8/edit?usp=sharing" },
      "schema:creator": [
    {
      "@type": "schema:Person",
      "@id": "https://orcid.org/0000-0001-7769-4272",
      "schema:name": "Remzi   elebi  "
    },
    {
      "@type": "schema:Person",
      "@id": "https://orcid.org/0000-0001-8381-1252",
      "schema:name": "Ilse van Lier"
    },
    {
      "@type": "schema:Person",
      "@id": "https://orcid.org/0000-0002-6500-4649",
      "schema:name": "Alie de Boer"
    },
    {
      "@type": "schema:Person",
      "@id": "https://orcid.org/0000-0003-4727-9435",
```

```
      "schema:name": "Michel Dumontier"
    }
    ],
      "schema:dateCreated": {
        "@value": "2021-12-01T00:00:00Z",
        "@type": "xsd:dateTime"
      },
      "schema:publisher": [
        { "@type": "schema:Organization",
    "@id": "https://www.maastrichtuniversity.nl/research/institute-data-science",
              "schema:name": "Institute of Data Science, Maastricht University"   }
      ],
      "schema:funding": [
        { "@id": "https://www.ufl.nl/en/" },
        { "@id": "https://www.limburg.nl/" }
      ],
      "schema:dateModified": {
        "@value": "2023-07-01T00:00:00Z",
        "@type": "xsd:dateTime"
      },
      "schema:datePublished": {
        "@value": "2023-07-01T00:00:00Z",
        "@type": "xsd:dateTime"
      },
 "void:vocabulary": [
    { "@id": "http://www.w3id.org/foodhkg/" },
    { "@id": "http://www.w3id.org/foodhkg/linksets/foodb" },
    { "@id": "http://www.w3id.org/foodhkg/linksets/snomedct" },
    { "@id": "http://www.w3id.org/foodhkg/linksets/go" },
    { "@id": "https://foodkg.org/linksets/meddra" }
  ],

      "schema:contactPoint": {
        "@type": "schema:Person",
        "schema:name": "Remzi Celebi",
        "schema:email": "remzi.celebi@maastrichtuniversity.nl"
      },
      "schema:subjectOf": { "@id": "https://ceur-ws.org/Vol-3890/paper-24.pdf" },
      "schema:image": [
      {
    "@type": "schema:ImageObject",
    "@id": "https://raw.githubusercontent.com/MaastrichtU-IDS/food-claims-kg/master/SIO_for_food_health_claim.png",
    "contentUrl": "https://raw.githubusercontent.com/MaastrichtU-IDS/food-claims-kg/master/SIO_for_food_health_claim.png"
       }
      ],
"schema:distribution": [
  {
    "@type": "schema:DataDownload",
    "schema:contentUrl": {
      "@id": "https://github.com/MaastrichtU-IDS/food-claims-kg/blob/master/output/food_health_kg.ttl"
    },
    "schema:name": "Food Health KG (TTL)",
```

```
      "schema:description": "RDF serialization of the Food Health Claims KG in Turtle format",
      "schema:encodingFormat": "text/turtle"
    },
    {
      "@type": "schema:DataDownload",
      "schema:contentUrl": {
        "@id": "https://github.com/MaastrichtU-IDS/food-claims-kg/blob/master/output/catalog-v001.xml"
      },
      "schema:name": "Catalog Metadata",
      "schema:description": "Catalog metadata in XML",
      "schema:encodingFormat": "application/xml"
    },
    {
      "@type": "schema:DataDownload",
      "schema:contentUrl": {
        "@id": "https://github.com/MaastrichtU-IDS/food-claims-kg/blob/master/output/food-claims-kg.xlsx"
      },
      "schema:name": "Food Health KG Spreadsheet",
      "schema:description": "Spreadsheet version of the Food Health Claims KG",
      "schema:encodingFormat": "application/vnd.openxmlformats-officedocument.spreadsheetml.sheet"
    }
        ],
        "schema:mainEntityOfPage": {
          "@type": "schema:WebAPI",
          "@id": "http://grlc.io/api-git/MaastrichtU-IDS/food-claims-kg"
        },
        "schema:contentUrl": {
          "@id": "https://graphdb.dumontierlab.com/repositories/FoodHealthClaimsKG"
        },
        "schema:version": "1.0",
        "schema:license": { "@id": "https://opensource.org/licenses/MIT" },

        "schema:keywords": "personalised nutrition, food health claim, knowledge graphs, FAIR data",
        "schema:publication": [
          { "@id": "https://ceur-ws.org/Vol-3890/paper-24.pdf" }
        ],
        "schema:inLanguage": [ "en" ],
        "void:triples": 7842,
        "dct:conformsTo": { "@id": "https://www.w3.org/TR/rdf11-concepts/" },
        "mod:sampleQueries": { "@id": "http://grlc.io/api-git/MaastrichtU-IDS/food-claims-kg" },
        "void:uriRegexPattern": "http://www.w3id.org/foodhkg/Instances/{id}",
       "void:exampleResource": {
       "dct:title": "Vitamin C Psychological Function Claim",
       "dct:description": "Example claim stating that vitamin C contributes to normal psychological function.",
       "dcat:mediaType": {
      "@value": "text/turtle",
```

```
    "@type": "xsd:string"
  },
  "adms:status": {
    "@value": "stable",
    "@type": "xsd:string"
  },
  "dcat:accessURL": {
    "@id": "http://www.w3id.org/foodhkg/Instances/XU1PVZE8962HgzcThlA8FBKmhdiJjsVm3jnDU3B12KE"
  }
},
      "dct:accessRights": { "@id": "http://publications.europa.eu/resource/authority/access-right/OPEN" },
      "prov:hadPrimarySource": {
        "@id": "https://ec.europa.eu/food/food-feed-portal/screen/health-claims/eu-register"
      },
      "void:uriSpace": "http://www.w3id.org/foodhkg/Instances/"
    },
{
  "@id": "https://foodkg.org/linksets/meddra",
  "@type": "void:Linkset",
  "void:subjectsTarget": { "@id": "http://www.w3id.org/foodhkg" },
  "void:target": { "@id": "http://purl.bioontology.org/ontology/MEDDRA" },
  "void:triples": 50
},
{
  "@id": "https://foodkg.org/linksets/go",
  "@type": "void:Linkset",
  "void:subjectsTarget": { "@id": "http://www.w3id.org/foodhkg" },
  "void:target": { "@id": "http://www.informatics.jax.org/vocab/gene_ontology" },
  "void:triples": 100
},
{
  "@id": "https://foodkg.org/linksets/snomedct",
  "@type": "void:Linkset",
  "void:subjectsTarget": { "@id": "http://www.w3id.org/foodhkg" },
  "void:target": { "@id": "http://purl.bioontology.org/ontology/SNOMEDCT" },
  "void:triples": 200
},
{
  "@id": "https://foodkg.org/linksets/foodb",
  "@type": "void:Linkset",
  "void:subjectsTarget": { "@id": "http://www.w3id.org/foodhkg" },
  "void:target": { "@id": "https://foodb.ca" },
  "void:triples": 1000
}

  ]
}
  </script>
</head>
<body>
  <noscript>This page contains structured metadata for a Knowledge graph.</noscript>
</body>
```

```
</html>
```

Listing 3: index.html containing schema.org metadata for food Health claims KGs