

# Part 4.8

## Metadata

- 1 A Methodology for Data Reuse
- 2 Phase 1 - Purpose Definition
- 3 Phase 2 - Information Gathering
- 4 Phase 3 - Language Definition
- 5 Phase 4 - Knowledge Definition
- 6 Phase 5 - Data Definition
- 7 Knowledge Graph Evaluation
- 8 Metadata**

## Metadata

- Sharing Resources
- Metadata definition

## KGE Projects

- An iTelos project, driven by a specific purpose, is a cooperative, composite project comprised of:
  - A *community of researchers and participants* including:
    - 1 producers, who are interested in *sharing* the different resources generated by a iTelos project for potential reuse.
    - 2 consumers, who are interested in *reusing* the different resources generated by an already existing iTelos project.
    - 3 intermediaries, who *generate* purpose-specific reusable resources to reduce language, knowledge and data heterogeneity between producers and consumers.
  - and resources ...

## DataScientia Community (Click Here!)



# MAKE THE DIFFERENCE WITH DATA

### Tweets from @DatascientiaF

 **DataScientia...** @Da... · Oct 17

Announcing the return of the @LogimSchool Summer School, taking place on July 8-12. Applications for mentors to guide groups of ~5 graduate students on a week-long project of the mentor's choosing to be submitted before 17/11/2023 here:

 docs.go...  
LOGML  
2024 ...

👍 ❤️ 📄

 **DataScientia...** @Da... · Sep 28

## DataScientia Community

### Who are the members of our community?

Anybody can be a member of the DataScientia Community. You do not need any specific skills to become a member. The only requirement is that you must be willing to learn about data, AI, and their impact on society.

You can decide the amount of effort devoted as a function of how much and how fast you want to learn.

### How can you contribute?

There are at least four different types of contributions. You can choose the one or more which best fit your competence, skills, and interests.

- The first type of contribution is to support the collection of person-centric data, as needed in one of the many experiments that we set up.
- The second type, if you have programming skills, is to participate to the open data and open-source community, in charge of development of the AI tools and software which allow DataScientia and its partners to operate.



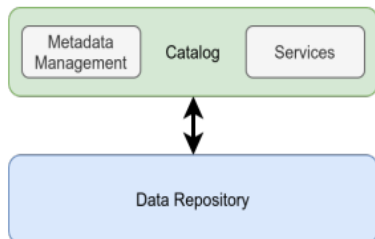
### What is a citizen science community?

A Citizen Science Community consists of people who voluntarily help conduct scientific research. Citizen scientists may be involved in various aspects of the research process, e.g., design experiments, collect data, analyze results, and solve problems.

## Project Resources

- The various resources generated by a iTelos project are stratified into:
  - *language resources*, e.g., domain language annotation spreadsheet.
  - *knowledge resources*, e.g., teleologies, teleontologies.
  - *data resources*, e.g., datasets, KGs.

## Sharing Resources - Catalog




- A web-base unique access point for data repositories
- Smart search and easier navigation for datasets
- Catalog only host dataset metadata


## Data Intermediary Catalogs

- The iTelos resources are organized and indexed in their respective (data) catalogs, namely:
  - 1 LiveLanguage Catalog
  - 2 LiveKnowledge Catalog
  - 3 LiveData Catalog
- The KGE Projects are also organized and indexed in a dedicated catalog: [LiveData Catalog](#)

# LiveLanguage: Example


About Us Datasets Organizations Services FAQ

🏠 / Datasets / Italian UKC Lexicon



**DataScientia**  
Unitas per Varietatem

---

**DataScientia  
Foundation**

---

DataScientia - a soon-to-be-established not-for-profit organization whose ultimate aim is the creation of a grass-roots community centered around the development and dissemination of a unitary knowledge-driven understanding of the people's diversity, as it is represented by the data in the Internet. DataScientia is being nurtured by the University of Trento, Department of Information Engineering and Computer Science.

## Italian UKC Lexicon

Italian is a language from the Indo-European family, spoken in Eurasia. The UKC Lexicon of Italian is represented as a lexico-semantic network. It consists of words, word senses, synsets, as well as sense-level and synset-level relationships.

### Resources

- Italian UKC Lexicon LMF format xml [\(Details\)](#)

<b>License</b>	Creative Commons Attribution NonCommercial ShareAlike
<b>Last updated</b>	2023-03-28
<b>Created</b>	2023-03-28
<b>Size (Bytes)</b>	2580172
<b>Release Date</b>	2023-03-28
<b>Distribution document or page</b>	
<b>Language</b>	Italian
<b>ISO-3 Language Code</b>	ita

⏪ ⏩ ⏴ ⏵ ⏶ ⏷ ⏸ ⏹ ⏺ ⏻ ⏼ ⏽ ⏾ ⏿



## LiveKnowledge: Example



[About Us](#) [Datasets](#) [Organizations](#) [Services](#) [FAQ](#)

🏠 / [Datasets](#) / OSM Lightweight Ontology



DataScientia  
Unitas per Varietatem

### DataScientia Foundation

DataScientia - a soon-to-be-established not-for-profit organization whose ultimate aim is the creation of a grass-roots community centered around the development and dissemination of a unitary knowledge-driven understanding of the people's diversity, as it is represented by the data in the Internet. DataScientia is being nurtured by the University of Trento, Department of Information Engineering and Computer Science, in collaboration

## OSM Lightweight Ontology

A lightweight ontology developed based on data from Open Street Maps.

### Resources

- [OSM-LO.UAN.owl](#) 

(Details)

A OWL RDF/XML distribution of the lightweight ontology developed based on data from Open Street Maps.

### Metadata

<b>Distribution</b>	osm-lwo-owl
<b>Keyword</b>	Geography
<b>Publisher</b>	DataScientia Foundation
<b>Category</b>	Society&Territory
<b>VersionNotes</b>	Version 1.0 - Unannotated.
<b>LandingPage</b>	Unknown.
<b>AccessRights</b>	Public

# LiveData: Example


[About Us](#) [Datasets](#) [Organizations](#) [Services](#) [FAQ](#)
[Home](#) / [Datasets](#) / [LiveData Trentino](#)


## DataScientia Foundation

DataScientia - a soon-to-be-established not-for-profit organization whose ultimate aim is the creation of a grass-roots community centered around the development and dissemination of a unitary knowledge-driven understanding of the people's diversity, as it is represented by the data in the Internet. DataScientia is being nurtured by the University of Trento, Department of Information Engineering and Computer

## LiveData Trentino

The LiveData Trentino catalog collects and shares datasets about the Trentino region (autonomous province of Trentino, Italy). The catalog provides data about the Trentino geography, as well as data about Trentino's society, like transportation, points of interest, facilities, and others. The LiveData Trentino catalog is part of the DataScientia Open Data Space. The data shared by the catalog are compliant with the quality criteria defined by DataScientia, moreover they are highly reusable and composable with other data produced following the DataScientia guidelines (more in detail, adopting the iTelos methodology).

### Resources

- [LiveData Trentino catalog](#) html

### Additional Info

<b>License</b>	Creative Commons Attribution
<b>Category</b>	Society and Territory
<b>Maintainer</b>	Simone Bocca
<b>Maintainer Email</b>	simone.bocca@unitn.it
<b>Keyword</b>	Space, Geography, Trentino

## KGE Catalog: Example



[About Us](#) [Datasets](#) [Organizations](#) [Services](#) [FAQ](#)

🏠 / [Datasets](#) / [Education in Trentino](#)



### KnowDive

UniTn DISI dep computer science research group

[🔗 Open in GitHub](#)

## Education in Trentino

This project was developed by Samuele Bortolotti and Erich Robbi for the Knowledge Graph Engineering course of the master's degree in Computer Science at the University of Trento.

### Resources

- [KGE - Education in Trentino](#) [html](#)

### Additional Info

<b>License</b>	Open Data Commons Attribution License
<b>Category</b>	Digital University
<b>Maintainer</b>	Simone Bocca
<b>Maintainer Email</b>	simone.bocca@unitn.it
<b>Author(s)</b>	Erich Robbi Samuele Bortolotti

## Metadata

- Sharing Resources
- Metadata definition

## Metadata - Definition

- Metadata is “*structured information that describes, explains, locates or otherwise makes it easier to retrieve, use or manage an information resource*” [NISO, 2017]
- Metadata, in general, has three main purposes:
  - 1 facilitate description of information resources
  - 2 facilitate organization of information resources
  - 3 facilitate discovery of information resources
- In the context of the iTelos Projects, the information resources include language, knowledge and data resources as stated earlier.

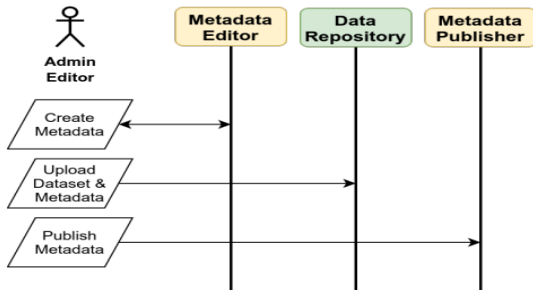
## Metadata - Scope

- In the context of *iTelos*, the scope of metadata is important with respect to two dimensions - *Quality* and *Reusability*
- Firstly, metadata allows the user to determine *data quality and fitness* for their DI project by helping them assess the usefulness of a data resource or a teleology relative to their requirement specification.
- Secondly, "*iTelos assumes the existence of a repertoire of teleologies and provides a rich set of metadata for reusing them*" (Giunchiglia *et al.*, 2021)
- It is essential to always adhere to a *metadata standard* for ensuring *reusability* and *shareability* of data and knowledge resources.

## Metadata and Catalogs

### Catalog - Services (1/2)

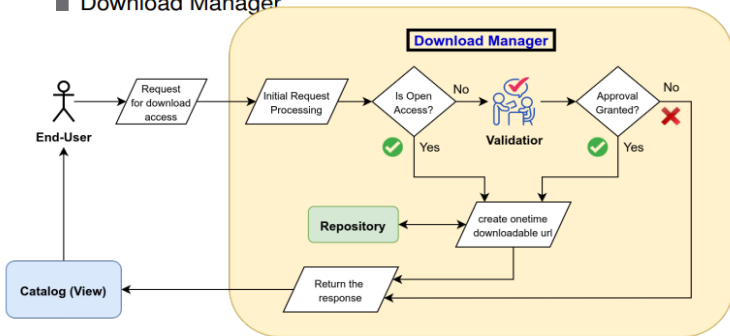
- Metadata Editor
- Metadata Publisher



## Metadata and Catalogs (Contd.)

### Catalog - Services (2/2)

#### ■ Download Manager





## iTelos Metadata Organization

- The iTelos Metadata schema is organized into the following layers:
  - 1 **People Metadata:** metadata attributes about individuals involved in a iTelos project.
  - 2 **Project Metadata:** metadata attributes about a iTelos project.
  - 3 **Dataset Metadata:** metadata attributes about dataset resources (whether language, knowledge or data resources) involved in a iTelos project.
- **"NR"** for a metadata attribute means it is non-repeatable, Information for that attribute should be recorded only once.
- iTelos metadata is recorded via **spreadsheets** where column headings are the individual metadata attributes (see next slides).

## People Metadata

- (a) ds:comIdentifier: this attribute encodes the DataScientia community identifier uniquely identifying a person within the DataScientia ecosystem. **NR**
- (b) ds:firstName: this attribute encodes the first name of the person in a natural language.
- (c) ds:lastName: this attribute encodes the last name of the person in a natural language.
- (d) ds:email: this attribute encodes the email id of the person.
- (e) ds:nationality: this attribute encodes the nationality of the person.
- (f) ds:gender: this attribute encodes the gender of the person. **NR**
- (g) ds:affiliation: this attribute encodes the organization to which the person is affiliated in a natural language.
- (h) ds:personalWebpage: this attribute encodes the URL of the personal webpage of the person. **NR**

## Project Metadata

1. ds:prjTitle: this attribute encodes the name of the DataScientia project in a natural language as a string.
2. ds:prjURL: this attribute encodes the dereferenciable URL of the DataScientia project. **NR**
3. ds:prjKeywords: this attribute encodes the various keywords in a natural language that can be utilized to quickly understand the theme of the project.
4. ds:prjType: this attribute encodes the type of the DataScientia project. e.g., Knowledge Resource Generation, Knowledge Resource Annotation, etc. **NR**
5. ds:prjDescription: this attribute can be used to provide a description of the DataScientia project in a natural language.
6. ds:prjStartDate: this attribute encodes the date of the commencement of a DataScientia project. **NR**

## Project Metadata (Contd.)

7. ds:prjEndDate: this attribute encodes the date of conclusion of a DataScientia project. **NR**
8. ds:prjFundingAgency: this attribute encodes the name of the agency or institution funding a DataScientia project.
9. ds:prjInput: this (repeatable) attribute encodes the various inputs (e.g., datasets, specifications, etc.) with respect to a DataScientia project.
10. ds:prjOutput: this (repeatable) attribute encodes the various outputs (e.g., datasets, domain languages, etc.) with respect to a DataScientia project.
11. ds:prjCoordinator: this attribute encodes the name of the research coordinator in charge of a DataScientia project. **NR**
12. ds:prjObservations: this attribute can be used to record any observations about a DataScientia project in a natural language.

## Datasets Metadata

1. ds:DatLicense: this attribute encodes the license of the dataset, e.g., CC-BY-SA-4.0 **NR**
2. ds:DatURL: this attribute encodes the dereferenceable URL of the dataset. **NR**
3. ds:DatKeyword: this attribute encodes the keywords which can quickly convey the topic of the dataset.
4. ds:DatPublisher: this attribute encodes the publisher of the dataset.
5. ds:DatCreator: this attribute encodes the creator of the dataset.
6. ds:DatOwner: this attribute encodes the owner of the dataset.
7. ds:DatLanguage: this attribute encodes the natural language(s) in which the dataset information is represented.
8. ds:DatLevel: this attribute encodes the knowledge level of the dataset, e.g., L1-2, L4. **NR**

## Datasets Metadata (Contd.)

9. ds:DatSize: this attribute encodes the byte size of the dataset. **NR**
10. ds:DatName: this attribute encodes the name of the dataset in a natural language.
11. ds:DatPublicationTimestamp: this attribute encodes the timestamp of the publication of the dataset in the respective catalog. **NR**
12. ds:DatDescription: this attribute encodes the description about the dataset in a natural language.
13. ds:DatVersion: this attribute encodes the version of the dataset. **NR**
14. ds:DatDomain: this attribute encodes the domain to which the dataset belongs, e.g., society and territory.
15. ds:DatFileFormat: this attribute encodes the file format of the dataset, e.g., OWL RDF/XML, Excel. **NR**

## Projects Metadata Sheet - Example

	A	B	C	D	E	F	G
1	Title	URL	Project Type	Description	Start Date	End Date	Funding Agency
2							
3	OSM General Lightweight Ontology General To Be Added.		Knowledge Resource Genera	This project focused on the gi	11-03-2023	11-04-2023	DataScientia Foundation
4	OSM General Lightweight Ontology Annotat To Be Added.		Knowledge Resource Annotat	This project focused on the ai	07-06-2023	10-06-2023	DataScientia Foundation
5							
6							
7							
8							

## Dataset Metadata Sheet - Example

	Description	Title	Distribution	Keyword	Publisher	Theme	Version	Notes	Landing Page	Access Rights	Creator
2	The DBpedia on	The DBpedia Or	dbpedia-owl	Wikipedia, Multil	DBpedia Organi	Upper Level	new version 4.2-		<a href="https://dbpedia.c">https://dbpedia.c</a>	Public	DBpedia Organi
3	The Bibliographi	The Bibliographi	bibo-owl	Bibliography		Upper Level	As of today, the		<a href="http://purl.org/on">http://purl.org/on</a>	Public	Bruce D'Arcus, F
4	FOAF is a projec	Friend Of A Frie	foaf-owl	People	Dan Brickley	Upper Level	No updates sinc		<a href="http://www.foaf-">http://www.foaf-</a>	Public	Libby Miller, Dar
5	This is the encor	CIDOC Concept	crm-owl	Customer management		Upper Level	2016: Annual rev		<a href="https://cidoc-crm">https://cidoc-crm</a>	Public	FORTH-ICS
6	CAT is an RDF	Data Catalog Vo	dcat.owl	Catalogs	W3C Data Exch	Upper Level	(2020-11-01) Gh		<a href="http://www.w3.or">http://www.w3.or</a>	Public	Richard Cygan
7	The Geonames	The Geonames	geonames-owl	geography	<a href="http://www.geon">Geonames.org</a>	Upper Level	2020: Annual rev		<a href="http://www.geon">http://www.geon</a>	Public	Bernart Vatant
8	A Geographic	OGC GeoSPAR	GeoSPARQL-owl	Geometry	Open Geospatia	Upper Level	2016: Annual rev		<a href="http://www.open">http://www.open</a>	Public	Open GeoSpatia
9	A vocabulary fr	NeoGeo Spatial	geovocab-owl	Topology		Upper Level	2016: Annual rev		<a href="http://geovocab">http://geovocab</a>	Public	Juan Martín Sales Andrus
10	A specializatio	OGC Geometry	gml-owl	Geometry		Upper Level	2016: Annual rev		<a href="http://www.open">http://www.open</a>	Public	Open GeoSpatia
11	This ontology i	General Transit	gtfs-owl	Travel		Upper Level	2016: Fixed vers		<a href="https://raw.githu">https://raw.githu</a>	Public	Pieter Colpaert,
12	A vocabulary t	The Opening Ho	ical-owl	Time	W3C	Upper Level	2019: Annual rev		<a href="https://github.co">https://github.co</a>	Public	Pieter Colpaert
13	A specification	Simplified Featu	sf-owl	Geometry		Upper Level	2016: Fixed vers		<a href="http://www.open">http://www.open</a>	Public	Open Geospat
14	The Simple Kn	Simple Knowled	skos-owl	Concept scheme	W3C	Upper Level	2015: Annual rev		<a href="http://www.w3.or">http://www.w3.or</a>	Public	Alistair Miles, S
15	This ontology i	Sensor, Observ	sosa-owl	IoT, Environmen	W3C/OGC Spati	Upper Level			<a href="http://www.w3.or">http://www.w3.or</a>	Public	W3C/OGC Spati
16	This ontology c	Semantic Sens	ssno-owl	IoT	W3C	Upper Level			<a href="http://www.w3.or">http://www.w3.or</a>	Public	W3C/OGC Spati