

Research Infrastructures supporting FAIR and Open Data

Tools and Practices in the SSH: the Example of CLARIN



Giulia Pedonese
16 giugno 2025



Sources

This presentation is the result of adapting the following sources:

- van der Lek, I., Fišer, D., Frontini, F., & Pedonese, G. (2024, agosto 30). Introduzione ai Dati Linguistici: Standard e Archivi Digitali. Zenodo. <https://doi.org/10.5281/zenodo.13911935>
- van der Lek, I., Fišer, D., Samardzic, T., Simonovic, M., Assimakopoulos, S., Bernardini, S., Milicevic Petrovic, M., & Puskas, G. (2023). Integrating research infrastructures into teaching: Recommendations and best practices (Versione 2) <https://doi.org/10.5281/zenodo.8114407>. [CC BY 4.0](#).
- CLARIN ERIC Official Website: <https://www.clarin.eu/> . [CC BY 2.0](#)

In this presentation

- Why Apply Open Science Best Practices
- What are the FAIR principles for
- Language Resources and Language Technologies in CLARIN
- What is a repository and how does it work
- What is a Research Infrastructure? The example of CLARIN ERIC
- Discovering published resources: the Virtual Language Observatory
- How to apply a user license
- How to properly cite a resource: the TASL scheme
- What is a Data Management Plan and why it is useful for research

Ice breaker



Open Science: a new paradigm

Open Science is a new approach to the scientific process based on cooperative work and new ways of disseminating knowledge through the use of digital technologies and new collaboration tools (European Commission, 2016)

Open science includes free access to scientific articles, access to public research data and collaborative research enabled by ICT tools and incentives (OECD, 2015)

<https://open-science.it/cos-e-l-open-science>

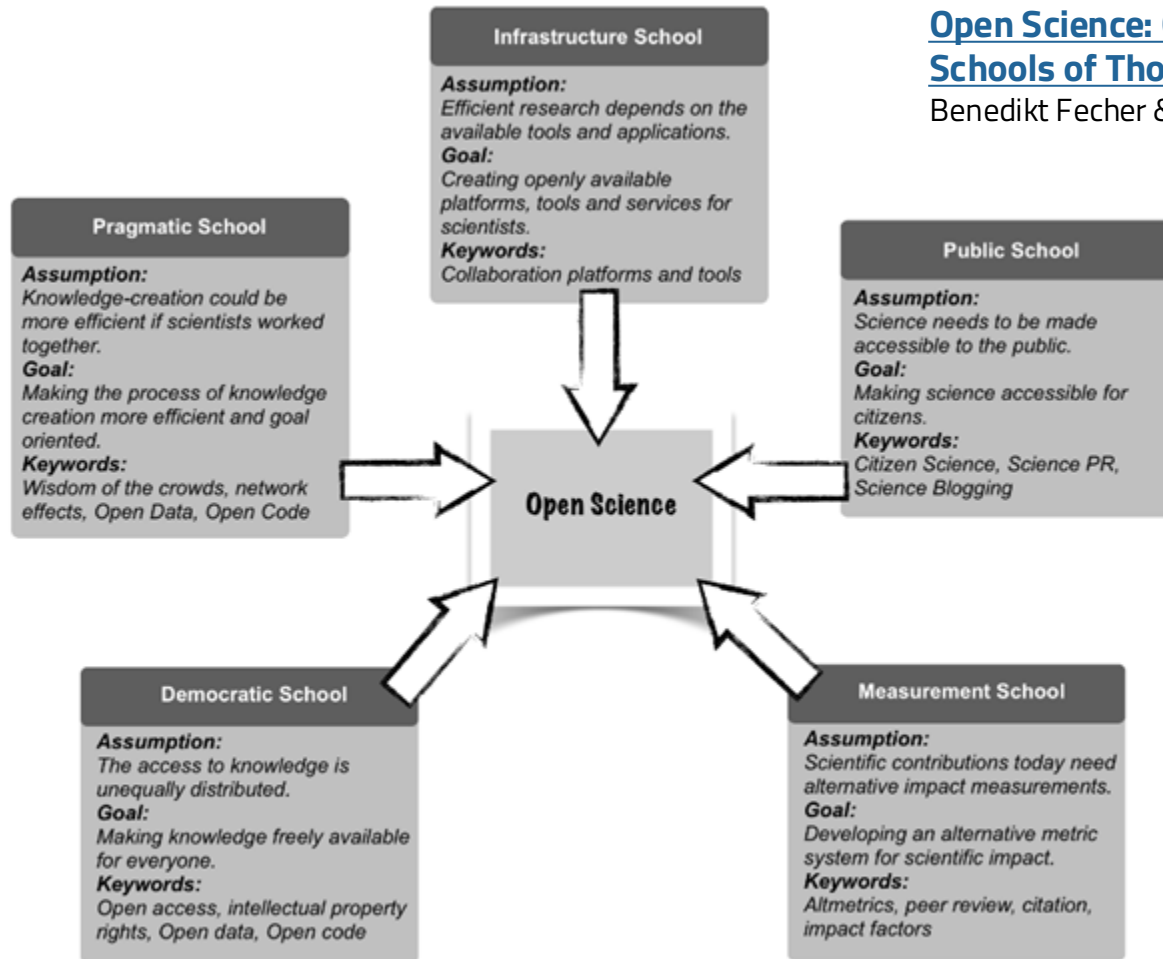
Research Data Lifecycle



1. Plan
2. Collection
3. Processing
4. Analysis
5. Archiving
6. Sharing
7. Reuse

Open Science: One Term, Five Schools of Thought

Benedikt Fecher & Sascha Friesike



FAIR principles

The FAIR Principles for the Management and Preservation of Scientific Data were published in 2016 with the main goal of providing guidelines to improve the management of digital objects in the Semantic Web. FAIR is an acronym for:

- **Findability** - Trovabilità
- **Accessibility** - Accessibilità
- **Interoperability** - Interoperabilità
- **Riusability** - Riusabilità

FAIR data management

'FAIR Guiding Principles for scientific data management and stewardship', 2016

<https://www.go-fair.org/fair-principles/>

FAIR Principles

Compliance



Findability

Resource and its metadata are easy to find by both, humans and computer systems. Basic machine readable descriptive metadata allows the discovery of interesting data sets and services.

- ✓ F1. Resource is uploaded to a public repository.
- ✓ F2. Metadata are assigned a globally unique and persistent identifier.



Accessibility

Resource and metadata are stored for the long term such that they can be easily accessed and downloaded or locally used by humans and ideally also machines using standard communication protocols.

- ✓ A1. Resource is accessible for download or manipulation by humans and is ideally also machine readable.
- ✓ A2. Publications and data repositories have contingency plans to assure that metadata remain accessible, even when the resource or the repository are no longer available.



Interoperability

Metadata should be ready to be exchanged, interpreted and combined in a (semi)automated way with other data sets by humans as well as computer systems.

- ✓ I1. Resource is uploaded to a repository that is interoperable with other platforms.
- ✓ I2. Repository meta- data schema maps to or implements the CG Core metadata schema.
- ✓ I3. Metadata use standard vocabularies and/or ontologies.



Reusability

Data and metadata are sufficiently well-described to allow data to be reused in future research, allowing for integration with other compatible data sources. Proper citation must be facilitated, and the conditions under which the data can be used should be clear to machines and humans.

- ✓ R1. Metadata are released with a clear and accessible usage license.
- ✓ R2. Metadata about data and datasets are richly described with a plurality of accurate and relevant attributes.

Example of data management: Linguistic Resources in CLARIN



What are language resource?

Language Resources are defined as the set of data and metadata relating to language in machine-readable form. They are used to build, improve or evaluate algorithms or systems of natural language (human language) and speech and, increasingly, for machine learning. They are also used as fundamental resources for, for example, linguistic and translation studies, for electronic publishing, for international transactions, for subject area specialists and for end users.

Source: [ELRA/ELDA](#)



Types of Language Resources

Language or linguistic resources are typically divided into categories according to the type of content that include:

Textual resources > written and spoken corpora

Lexical resources > lexicons, dictionaries and terminology databases

NLP tools > lemmatisers, PoS taggers, parsers etc.

Linguistic resources include not only the data, but also their description, or metadata, which enrich the data with additional information, such as the structural division into books or linguistic features such as PoS tags and syntactic functions

Why are they useful?

Language resources (LR) are the raw material for the development and upgrade of language technologies (LT), the medium for transmitting information and knowledge (if possible in the most efficient and effective way), the content for developing culture and civilising societies.

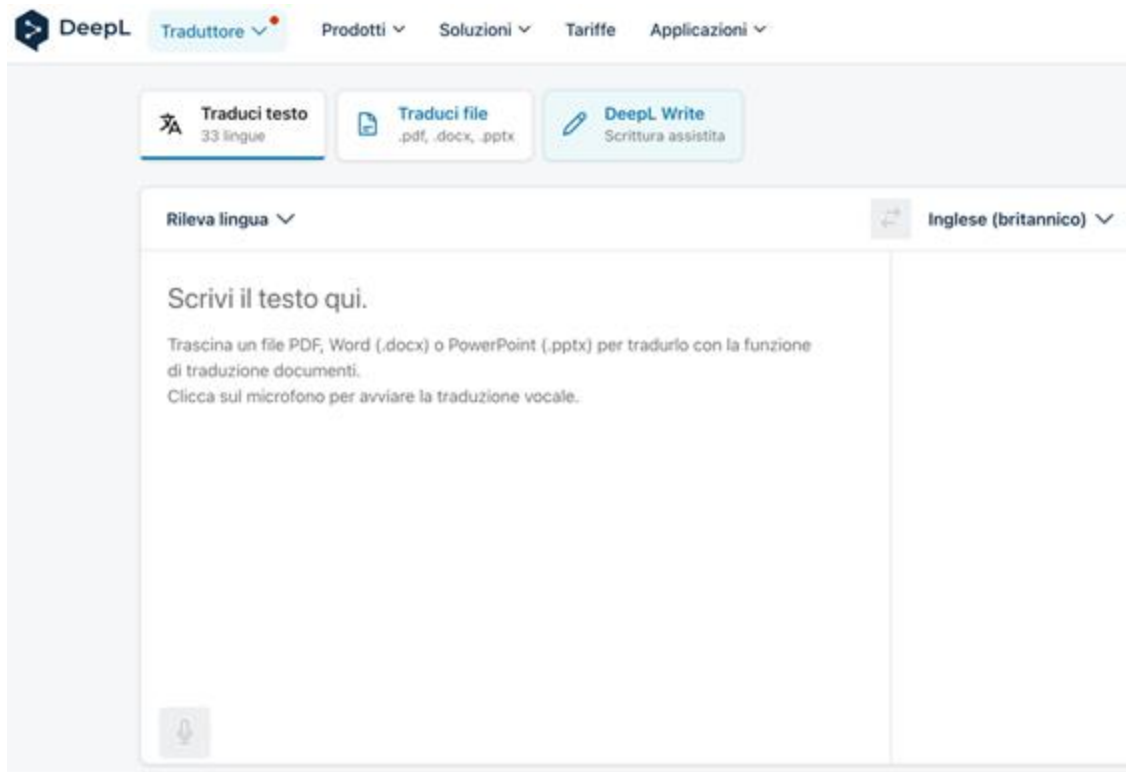
LR, in combination with LT, have greatly changed the user experience and the interactive possibilities of apps, tools and systems and public media over the years. Therefore, LR - in combination with LT - have a huge economic and social impact.

Source: [LT Innovate](#)

Language Technologies

Language technologies are computational methods, computer programs and specialized electronic devices for analyzing, producing or modifying texts and speech, e.g. spell checkers, machine translation services, speech synthesis, translation memories, computer-assisted translation, etc.

To know more: [Nimdzi Language Technology Atlas](#)

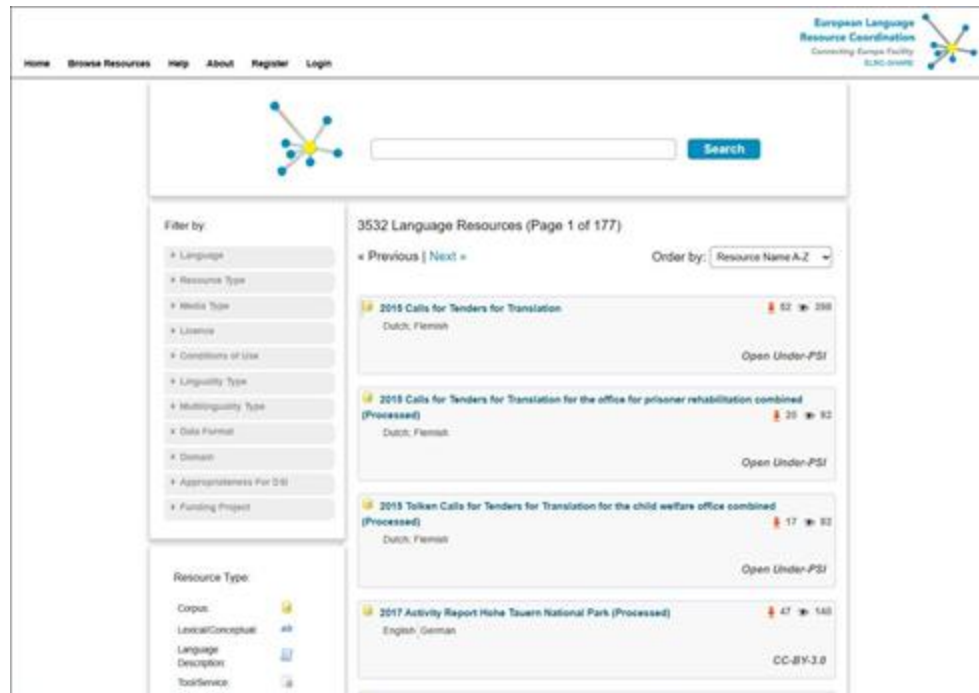


Where to find linguistic resources: repositories

“We define a repository – or digital archive or repository – as a service managed by research organizations, where research materials are preserved, managed and made accessible.”

Source: [DataCite](#)

To view the different types of linguistic resources and understand how they are described (the metadata model), you can consult the page [ELRC-SHARE repository](#)



ILC4CLARIN: finding published resources



<https://ilc4clarin.ilc.cnr.it/>

Linguistic Resources in ILC4CLARIN

To see the different types of Language Resources and understand the metadata model, see for example the ILC4CLARIN repository

Showing 1 through 5 out of 5 results

1 

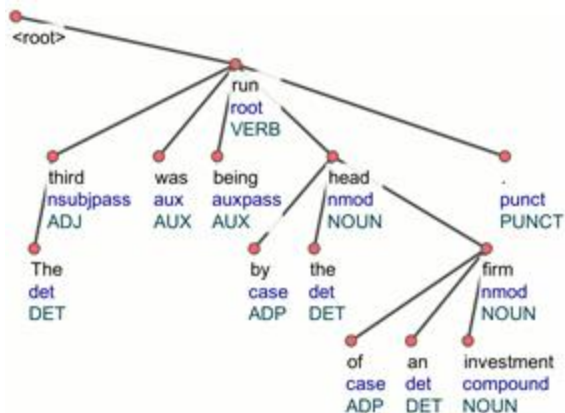
| Type | |
|---------------------------|-----|
| corpus | 443 |
| languageDescription | 1 |
| lexicalConceptualResource | 59 |
| text | 293 |
| toolService | 17 |

Example of a corpus in ILC4CLARIN

The screenshot shows the ILC4CLARIN Repository Home page for the item 'Travel Writings on Italy'. The page includes a header with the repository name and a 'View Item' link. Below the header, there is a yellow banner with a citation instruction and a 'HANDLE' button (annotated with a red circle 1). The main content area lists various metadata fields: Authors (Sprungov, Rachele, annotated with a red circle 2), Item identifier (http://hdl.handle.net/20.500.11752/OPEN.976), Project URL, Demo URL, Referenced by (https://ojs.ub.uni-erlangen.de/handle/10.1515/erlangen-2018-0001, annotated with a red circle 3), Date issued (2018), Type (corpus, text), Size (50 files), Language(s) (English), Description (A collection of travel writings - non-fictional narratives (reports, diaries, letters) and guidebooks - about Italy written by English native authors and published between the country unification and the beginning of the 20th century. Texts are available in both PDF and TEI-XML format. Texts have been collected from Project Gutenberg. TEI was automatically generated by GutenTag v0.1.4 and then manually checked and integrated, annotated with a red circle 4), Publisher (Fondazione Bruno Zevi - Trento, annotated with a red circle 5), Subject(s) (Travel writings, Historical texts), and Collection(s) (ILC4CLARIN - OPEN Data & Tools). At the bottom, there are two buttons: 'Download instructions for command line' and 'Download all files in item (10.57 MB)', with the latter annotated with a red circle 6.

1. Citation information via handle
2. Metadata fields describing the corpus
3. Corpus is cited in a journal
4. Corpus is described, e.g. texts are available in UTF-8 and TEI-XML
5. Publisher information
6. Download instructions

Linguistic Resources in CLARIN



CLARIN

Common Language Resources and
Technology Infrastructure

<https://www.clarin.eu/>

CLARIN Resource Families

Resource Families

- Well-curated corpora and lexical resources organized by data type and language
- NLP tools
- Licensing information
- Can be downloaded directly
- Some corpora are available through concordances, e.g. Korp, Corpuscle or KonText

CLARIN Resource Families

Corpora

- Computer-Mediated Communication Corpora
- Corpora of Academic Texts
- Historical Corpora
- L2 Learner Corpora
- Legal Corpora
- Literary Corpora
- Manually Annotated Corpora
- Multimodal Corpora
- Newspaper Corpora
- Oral History Corpora
- Parallel Corpora
- Parliamentary Corpora
- Reference Corpora
- Sign Language Resources
- Spoken Corpora

Lexical Resources

- Language Models
- Lexica
- Dictionaries
- Conceptual Resources
- Glossaries
- Wordlists

Tools

- Corpus Query Tools
- Normalisation
- Named Entity Recognition
- Part-of-Speech Tagging and Lemmatisation
- Tools for Sentiment Analysis

What is CLARIN

CLARIN stands for **Common Language Resources and Technology Infrastructure**



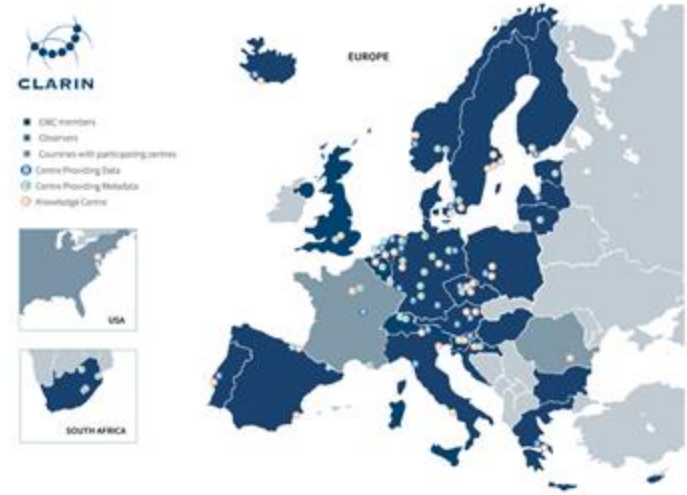
CLARIN is a distributed digital infrastructure that provides easy and sustainable access to a wide range of linguistic data and tools to support research in the humanities, social sciences and beyond. CLARIN provides access to multimodal digital linguistic data (text, audio, video) and advanced tools to explore, analyse or combine these data sets.

CLARIN...

- It belongs to the Social Sciences and Humanities cluster and is an integral part of the European Open Science Cloud
- It has had ESFRI ERIC status since 2012, Landmark status since 2016
- It acts as an ecosystem for knowledge sharing and training
- It provides easy and sustainable access to scholars in the humanities and social sciences and beyond, to digital linguistic data (in written, spoken or multimodal form); to advanced tools to discover, explore, exploit, annotate, analyse or combine them through single sign-on

CLARIN Today

- Una rete distribuita di 72 centri ([Centre Registry](#))
- 25 membri: AT, BE, BG, CY, CZ, DK, EE, ES, FI, GR, HR, HU, IS, IT, LT, LV, NL, NO, PL, PT, SE, SI, ZA, UK
- 1 osservatore: CH
- Terza parte: Università Carnegie Mellon, USA



CLARIN ERIC



Consorzi Nazionali



Centri CLARIN

National Consortia: CLARIN-IT

CLARIN-IT is the Italian node of CLARIN. Italy became the 16th full member of CLARIN ERIC in 2015. The founding member of the national Consortium is the Institute for Computational Linguistics "Antonio Zampolli" of the National Research Council. CLARIN is involved in the H2IOSC project through its Italian national consortium CLARIN-IT



CLARIN-IT

CLARIN-IT comprehends the following centres:

- [ILC4CLARIN](#) istituito presso il CNR Istituto di Linguistica Computazionale “Antonio Zampolli” di Pisa, è un centro di fornitura servizi ([Type B](#))
- [EURAC Research CLARIN Centre](#) (ERCC) presso l’EURAC Research Association di Bolzano, che è un centro di fornitura metadati ([Type C](#))
- [The Knowledge Centre for Computer-Mediated Communication and Social Media Corpora](#) (CKCMC), che è un centro CLARIN [di Tipo K](#), che mette a disposizione le proprie competenze e la propria conoscenza su aspetti del dominio disciplinare di CLARIN
- [The CLARIN Knowledge Centre for Digital and Public Textual Scholarship](#) (DiPText), anch’esso un centro di tipo K distribuito fra due partner: CNR-ILC e l’Università Ca’ Foscari di Venezia

FAIR principles in CLARIN

- FINDABLE through access points such as the Virtual Language Observatory and Federated Content Search
- ACCESSIBLE through CLARIN centers that provide repositories, metadata, and a federated login system
- INTEROPERABLE through the Language Switchboard, the use of vocabularies (Concept Registry), and clear and well-documented ways of linking metadata
- REUSABLE through licensing, formats, and standards, active curation of metadata, and link controls

CLARIN core services

- Repository services to ensure that language resources can be archived and made available to the community in a reliable manner and to help researchers maintain their resources sustainably
- [Virtual Language Observatory](#) provides an easy-to-use interface, which enables a process of searching and discovering resources from a wide variety of domains
- [Federated Content Search](#) a search engine that links to local data collections available in the centers.
- [Language Resource Switchboard](#) helps users find a web application to process their linguistic data
- [Virtual Collection Registry](#) provides a registry where users can create and publish their virtual collections

CLARIN Depositing Services

Many CLARIN centers offer a storage service. This provides many advantages:

- Long-term archiving and curation of language resources, data sets and tools
- Description of resources with specific metadata and assignment of persistent identifiers (PIDs), e.g. Handles, that allow easy search and citation
- Password-protected resources can be made available via an institutional login
- Once integrated into the CLARIN infrastructure, resources can be more easily analyzed and enriched with various language tools

Servizi di Deposito

Depositare una risorsa in CLARIN:

- Search for a repository hosted by a CLARIN Centre curating the type of data you produced [Depositing Services](#)
- Check for data compatibility in the [CLARIN Standards Information System](#)
- Contact the type B center hosting the repository for data preparation
- Follow the procedure described in the repository guidelines


[How to deposit](#)

How to access CLARIN

All users can freely explore CLARIN central services to search for language resources and expertise. Due to licensing restrictions, some resources are only available to academic users and login is required using institutional credentials or CLARIN credentials.

Academic users from all participating countries can access and use the language resources available in the CLARIN data centers with a single sign-on through the CLARIN Service Provider Federation using their institutional credentials.

CLARIN Service Provider Federation

 Sign in via the CLARIN Service Provider Federation



Select your home organisation below. This is usually the organisation where you work or study. Signing in here will allow you to access certain CLARIN resources and services which are only available to users who have logged in. If you cannot find your organisation in the list below, please select the clarin.eu website account and use your CLARIN website credentials. If you don't have such credentials you can register an account [here](#). For questions please contact spf@clarin.eu.

Warning: It appears as if you visited this page directly, this will not work. Please login [via the service](#) you are trying to access.

Home organisation list



All countries



IRCCS ISMETT - Palermo



Italy

ISMETT

University of Palermo



Italy



UNIVERSITÀ
DEGLI STUDI
DI PALERMO

CLARIN Identity Provider

CLARIN Identity Provider · Help



CLARIN account registration

Thank you for your interest in CLARIN. Please complete the form below.

After your registration is processed (normally within two working days), an automated email will be sent to your email address for verification. Click on the included link to verify your email address and activate your account. **(if you do not receive this email please make sure to check your spam folder).** After your account is active you can [download, explore and analyze password-protected language data with the CLARIN Identity Provider](#).

The minimal password length is 8 characters. Use a mix of letters, numbers and symbols.

Signing up with **an academic email address** will ensure maximal access rights and fast approval. We strongly recommend it.

For questions and remarks about this procedure, please check our [knowledge base](#) or contact accounts@clarin.eu

E-mail address: *

Trovare risorse linguistiche pubblicate

Virtual Language Observatory

- A catalog that collects metadata about language resources available in distributed repositories
- Does not contain language resources, but helps locate them through persistent identifiers
- Even if a resource has restricted access, metadata is always freely accessible
- Uses faceted search to narrow searches

NB: use of **licenses**



CLARIN Virtual Language Observatory

VLO / Faceted search / Search results

treasure island

Showing 3 results for **treasure island** x

Use the categories below to limit the search results to those matching the selected value(s).

Language

Collection

Resource type

File type

Temporal Coverage

Availability

Search options

Treasure Island / Robert Louis Stevenson

(Part of OTA Core Collection)

No description

English

The search results include 1 record with the same title.

Landing page for this record

Mind the Gap

(Part of DASH-DE Repository)

Mind the Gap was founded in 1988 by Tim Wheeler and Susan Brown and has since become one Europe's leading disability theatre companies, with operations in both the UK and internationally. Its primary aim of Mind the Gap's extensive team is to create equal opportunities for actors with learning disabilities within the...

English

La isla del tesoro - Spanish

(Part of DASH archive: Movima)

The Movima Archive Movima is a genetically unclassified language spoken in the so-called Moxos region in the savannahs of the Bolivian Amazon area. It is still spoken by more than 1,000 people (1,432 in 1996). Most speakers are over 50 years old and bilingual in Spanish. There are only very few children who

Treasure Island / Robert Louis Stevenson

Please use the following text to cite this item or export to a predefined format:

Stevenson, Robert Louis, 1850-1894, 1994, *Treasure Island / Robert Louis Stevenson*, Literary and Linguistic Data Service, <http://hdl.handle.net/20.500.14106/2055>.

Share: [f](#) [t](#) [w](#) [s](#)

Oxford Text Archive

Authors Stevenson, Robert Louis, 1850-1894

Date of publication 1882

Type Text

Language(s) English

OTA identifier ota:2055

Collection(s) OTA Core Collection

Show full item record

This item is **Publicly Available** and licensed under:
Attribution-NonCommercial-ShareAlike 3.0 Unported (CC BY-NC-SA 3.0)

Files for this item

| | |
|-------------|--|
| Name | treas10-2055.txt |
| Size | 370.51 KB |
| Format | Text file |
| Description | Version of the work in plain text format |

0/008

0/008

0

OXFORD
TEXT
ARCHIVE

FACULTY OF
LINGUISTICS,
PHONOLOGY
AND
PHONETICS

Browse

All of the Repository

My Account

Login

Statistics

Statistics

BETA

General Information

Deposit

Cite

Oxford University users

Submission Lifecycle

FAQ

About

Help Desk

Privacy policy

I Metadati

The purpose of metadata for a learning resource is to enable cataloging and discovery through standardized means to report:

- Who created the resource
- What the content of the resource is
- When the resource was created
- Where the resource is located
- Why the resource was created

Metadata must be in standard (i.e. compatible and shared with other resources) and machine-readable formats to be indexable, searchable, comparable, and storable

Persistent Identifiers (PIDs)

A persistent identifier (PID) is a type of metadata. Its purpose is to uniquely label a digital object and ensure that this label does not change over time

There are different types of PIDs, such as DOI (digital object identifier), Handle (HNDL), ARK (Archival Resources Key)

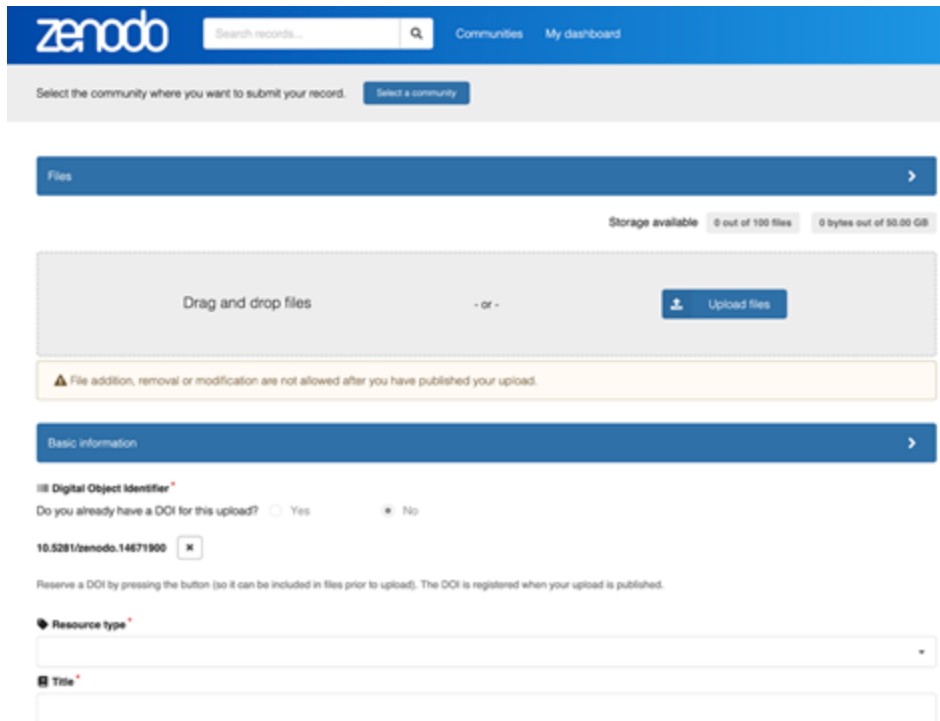
Examples:

- DOI: [10.5281/zenodo.8305539](https://doi.org/10.5281/zenodo.8305539)
- Handle: hdl.handle.net/20.500.11752/ILC-567

Persistent Identifiers (PIDs)

Using a PID you can get to an accessible page with the listed digital object and its metadata. Actual access to the digital object from this page can be restricted. PIDs are machine readable and help distinguish between:

- Different materials
- Different versions of the same material, for example:
<https://zenodo.org/records/13816183>



The screenshot shows the Zenodo website's upload interface. At the top is a blue header with the 'zenodo' logo, a search bar, and links for 'Communities' and 'My dashboard'. Below the header is a grey bar with the text 'Select the community where you want to submit your record.' and a 'Select a community' button. The main content area has a blue bar labeled 'Files' with a right arrow. Below this, a storage status bar shows 'Storage available' with '0 out of 100 files' and '0 bytes out of 50.00 GB'. The central upload area is a large grey box with the text 'Drag and drop files' and '- or -' followed by an 'Upload files' button with a cloud icon. Below this is a yellow warning box: '⚠ File addition, removal or modification are not allowed after you have published your upload.' The 'Basic information' section is a blue bar with a right arrow. It contains a 'Digital Object Identifier' section with the question 'Do you already have a DOI for this upload?' and radio buttons for 'Yes' and 'No'. Below this is a text input field containing '10.5281/zenodo.14671900' and a small 'x' icon. A note below states: 'Reserve a DOI by pressing the button (so it can be included in files prior to upload). The DOI is registered when your upload is published.' The 'Resource type' section has a dropdown menu, and the 'Title' section has a text input field.

Intellectual Property Rights

Intellectual property refers to the creation of intellectual assets. Intellectual property rights (IPR) protect the interests of creators and owners by providing them with rights to their creation.

For the purpose of research, copyright and related rights, the main branches of IPR, are used. They are defined to protect literary and artistic creations, performances and phonograms, defining the rights of authors, owners, performers, producers and broadcasters.

Note: Any work created by the user is by default the copyright holder for the user himself or his institution/employer.

Copyright

An original work may be protected by copyright law, which grants the owner the exclusive right to control certain rights such as reproduction. The copyright is held jointly by all authors, or it may be owned by the employing institution.

Use of copyrighted works requires permission from the owner. Permission may not be required for works licensed in the public domain and for uses covered by copyright exceptions



copyright

all rights reserved

Copyright

Only the copyright holder is allowed to distribute what is created. If you want to pass this right on to other people, you can do so through a license.

In the absence of exceptions or limitations, you can reuse an existing work if it is licensed to you or if it is licensed to the public through a public license such as Creative Commons (CC) or Free-Libre / Open Source Software (FLOSS) licenses.







How to apply a CC license?

The content of these slides is adapted from the following sources:






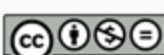
- CC FAQ website distribuito con licenza CC BY 4.0
- Creative Commons license Wikipage (16 gen 2025) su Wikipedia distribuito con licenza Creative Commons Attribution-ShareAlike 4.0
- CC License Compatibility Chart

How to apply a CC license?

A license defines the rights and obligations regarding the use of your work. One of the best sources of information for licensing reusable educational materials is Creative Commons. There are many CC license options available that are based on the four fundamental rights:

| Icon | Right | Description |
|---|-------------------------------------|---|
|  | Attribution (BY) | Licensees may copy, distribute, display, perform and make derivative works and remixes based on it only if they give the author or licensor the credits (attribution) in the manner specified by these. Since version 2.0, all Creative Commons licenses require attribution to the creator and include the BY element. |
|  | Share-alike (SA) | Licensees may distribute derivative works only under a license identical to ("not more restrictive than") the license that governs the original work. (See also copyleft .) Without share-alike, derivative works might be sublicensed with compatible but more restrictive license clauses, e.g. CC BY to CC BY-NC.) |
|  | Non-commercial (NC) | Licensees may copy, distribute, display, perform the work and make derivative works and remixes based on it only for non-commercial purposes. |
|  | No derivative works (ND) | Licensees may copy, distribute, display and perform only verbatim copies of the work, not derivative works and remixes based on it. Since version 4.0, derivative works are allowed but must not be shared. |

How to apply a CC license?

| License name ⇅ | Abbreviation ⇅ | Icon ⇅ | Attribution required ⇅ | Allows remix culture ⇅ | Allows commercial use ⇅ | Allows Free Cultural Works ⇅ |
|---|----------------|--|------------------------|------------------------|-------------------------|------------------------------|
| Attribution | BY |  | Yes | Yes | Yes | Yes |
| Attribution-ShareAlike | BY-SA |  | Yes | Yes | Yes | Yes |
| Attribution-NonCommercial | BY-NC |  | Yes | Yes | No | No |
| Attribution-NonCommercial-ShareAlike | BY-NC-SA |  | Yes | Yes | No | No |
| Attribution-NoDerivatives | BY-ND |  | Yes | No | Yes | No |
| Attribution-NonCommercial-NoDerivatives | BY-NC-ND |  | Yes | No | No | No |

[Creative Commons license Wikipage](#) (2025, Jan 14) su Wikipedia con licenza [Creative Commons Attribution-ShareAlike License 4.0](#) / immagine tagliata

Choosing a license

It is recommended to use the less restrictive CC BY 4.0 license, which only requires crediting the creator when reusing, when creating new teaching materials. The ND right is at the other extreme and is not recommended for licensing FAIR teaching materials, as it severely limits their reusability in adaptations.

The CC0 public domain license is also not recommended, as different rules apply in different countries.



Choosing a license

The wizard CC License Chooser helps authors share their work in a standardized way, providing copyright licenses that allow sharing and reusing the creative work under the conditions chosen.

To apply a license, simply indicate which CC license you intend to apply to your work. It is then necessary to check the terms of the chosen license and comply with the requirements contained therein.

Choosing a license

It is recommended to include a link to the relevant CC license act (for example <https://creativecommons.org/licenses/by/4.0>) or the full text of the license (in the README file).

Example «This work is licensed under [Creative Commons Attribution 4.0 International License](#).»

License compatibility

If my lesson is a combination of pre-existing material from different sources and reused as is, I just need to properly credit each part and state where it came from and what license it is under.

If you are adapting material that has previously been released under a CC license, the original CC license always applies to the material you are adapting, even after it has been adapted. The license you can choose for your reworking (called the “adapter license”) depends on which license applies to the original material.

NOTE: A modification reaches the level of an adaptation under copyright law when the modified work builds on the earlier work but exhibits enough new creativity to be protected by copyright.

| Adapter's license chart | | Adapter's license | | | | | | |
|-------------------------|----------|-------------------|-------|----------|----------|-------|-------|----|
| | | BY | BY-NC | BY-NC-ND | BY-NC-SA | BY-ND | BY-SA | PD |
| Status of original work | PD | | | | | | | |
| | BY | | | | | | | |
| | BY-NC | | | | | | | |
| | BY-NC-ND | | | | | | | |
| | BY-NC-SA | | | | | | | |
| | BY-ND | | | | | | | |
| | BY-SA | | | | | | | |

License compatibility

Note: CC does not recommend using a license if the corresponding box is yellow, even if this is technically allowed by the license terms. If you do so, you should be careful to indicate that the adaptation involves multiple copyrights with different terms, so that downstream users are aware of their obligations to respect the licenses of all rights holders.

Dark gray boxes indicate licenses that cannot be used as an adapter license.

Attribution

Attribution is a moral right of authors that protects the personal relationship between the author and the work created, even if the creator does not own the copyright. Recognition of reused materials through attribution is always necessary (even if it is not a requirement of the license, as in the case of public domain licenses). Ideal attribution follows the TASL approach

Title - what is the name of the work

Author - who allows the work to be used (name and link)

Source - where the work can be found (link added to the title)

License - how the work can be used (name and link to the license)

How to attribute correctly?

If your work is a modification or adaptation of another work, indicate that and provide attribution to the creator of the original work. You should also include a link to the work you modified and indicate which license applies to that work.

For works created by others and that you are incorporating into your own work, you should make it easy for others to know who created which parts of the work, using the TASL approach.

See also [Recommended practices for attribution](#)

Come Attribuire Correttamente?

If you reuse a complete resource, it is recommended that you provide attribution at the beginning of the unit. The same applies to reused parts of a lesson unit (text only or a combination of text and images).

Attribution for reused images and tables should be placed below the content.

<https://h2iosc-training-platform.ilc4clarin.ilc.cnr.it/>

Attribuzione

Questo corso traduce in italiano e aggiorna i materiali di:

van der Lek, Iulianna; Fišer, Darja. (2023). *Introduction to Language Data: Standards and Repositories*. In [UPSKILLS Learning Content](https://upskillsproject.eu/project/standards_repositories/). https://upskillsproject.eu/project/standards_repositories/, CC BY 4.0.

L'adattamento si è svolto nell'ambito del progetto [Humanities and cultural Heritage Italian Open Science Cloud](#) Work Package 8 "Training, Capacity Building, Engagement", a cura del personale CNR-ILC dedicato all'Attività 8.2 "Teach CLARIN, Teach with CLARIN".

Progetto H2IOSC - Humanities and cultural Heritage Italian Open Science Cloud finanziato dall'Unione Europea NextGenerationEU - PNRR M4C2 - Codice progetto IR0000029 - CUP B63C22000730005.



Immagine tagliata da: van der Lek, Iulianna, Darja Fišer, Francesca Frontini, e Giulia Pedonese. «Introduzione Ai Dati Linguistici: Standard E Archivi Digitali». 31 agosto 2024. <https://doi.org/10.5281/zenodo.13911935>, con licenza [CC BY 4.0](#).

Come Attribuire Correttamente?

To summarize:

- Pay attention to the ways indicated by the authors and give them precedence
- Check the README file or the presence of a .cff file
- If necessary, adapt the citation style to the bibliography you are composing
- If there is no indication, follow the TASL scheme

How to correctly attribute?

In order for others to cite our work properly, we recommend that a standard citation be established and explicitly communicated (at the bottom of the Syllabus, in the Introduction, in the Reuse Guide, in the .cff file for deposit, etc.).

Example:

Modalità di citazione

È possibile adattare e condividere il contenuto sotto la licenza [CC-BY 4.0](#) a condizione che:

- si dia il giusto credito agli autori utilizzando la **modalità di citazione standard**:

van der Lek, Iulianna, Darja Fišer, Francesca Frontini, e Giulia Pedonese. «Introduzione Ai Dati Linguistici: Standard E Archivi Digitali». 31 agosto 2024. <https://doi.org/10.5281/zenodo.13911935>.

- si includa il link alla licenza.
- si indichi se sono state apportate modifiche al contenuto.

Immagine tagliata da: van der Lek, Iulianna, Darja Fišer, Francesca Frontini, e Giulia Pedonese. «Introduzione Ai Dati Linguistici: Standard E Archivi Digitali». 31 agosto 2024. <https://doi.org/10.5281/zenodo.13911935>, con licenza [CC BY 4.0](#).

Quiz



What is a Data Management Plan?

A Data Management Plan

A **Data Management Plan (DMP)** is a formal document that describes how data will be managed during and after a research project.

Its main purpose is to **ensure that data are well-organized, accessible, usable, and preserved over time**, in line with FAIR (Findable, Accessible, Interoperable, Reusable) principles and open science best practices.

Key elements

A DMP usually includes information on:

Type of data

- What data will be collected, generated, or reused?
- In what format will it be stored?

Standards and metadata

- What standards will be followed to ensure interoperability and reuse?
- How will the data be documented?

Key elements

- **Storage and Security**

- Where and how will data be stored during the project?
- What security measures will be taken to protect them?

- **Sharing and access**

- What data will be made publicly accessible?
- Will there be specific restrictions or licensing requirements?

Key elements

- **Responsibility and long-term management**
 - Who will be responsible for data management?
 - Where and for how long will they be stored after the end of the project?

Why having a DMP is important?

- Helps researchers organize and manage data efficiently
- Facilitates sharing and reuse by other scholars
- Is often a requirement for obtaining funding from public and private entities
- Reduces the risk of data loss or corruption over time

Some funding agencies, such as the European Commission and the National Science Foundation (NSF), require a DMP as part of project proposals.

Plan and follow your data

Create machine actionable DMPs.

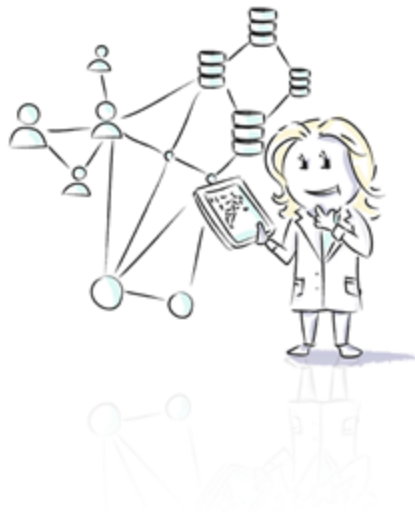
Configure to best fit your discipline.

Link to EOSC components out of the box.

Share easily in your repository.

Bring your Data Management Plans closer to where data are generated, analysed and stored.

Start your DMP



<https://argos.openaire.eu/splash/>

Exercise

Guided tour of a DMP with a case study

<https://docs.google.com/document/d/1p38h1jUdP8iFT-lpc9iR2MNxa44NP9j-NEoz5x3Rb2s/edit?tab=t.0>

Workshop

Try and fill in a DMP with the template

Register now!



What is the H2IOSC Training Environment?

The H2IOSC Training Environment is a training management platform designed to **offer a highly interactive virtual learning environment**. It ensures a smooth and inclusive experience for all users through:

- ✓ A user-friendly interface
- ✓ Access from desktop and mobile
- ✓ No time or space limits

The platform emphasizes interactivity by offering quizzes, simulations, hands-on exercises and, gamification tools that make learning more engaging.

Questionnaire



Iniziative della Comunità CLARIN

Call for H2IOSC Transnational Mobility Grants

[Home](#) / [News](#) / Call for H2IOSC Transna...

The H2IOSC project is pleased to announce a new opportunity for researchers: the Transnational Mobility Grants. These grants are designed to **encourage research collaborations and knowledge exchange between CLARIN centres in Italy and across Europe.**

Grant Opportunities

Long Mobility Grants: 5 grants available for research visits of 2 weeks, each funded between €6,000 – €7,000.

Short Mobility Grants: 11 grants available for visits of up to 1 week, each funded between €3,000 – €4,000.



<https://www.h2iosc.cnr.it/call-for-transnational-mobility-grants/>

CLARIN Community Initiatives

[CLARIN Newsflash](#)

[CLARIN Café](#)

Thank you!

francesca.frontini@cnr.it

giulia.pedonese@cnr.it

