CLARIN-IT

the Italian Common Language Resources and Technology Infrastructure



CLARIN-ITIFY

... tutto quello che avreste voluto sapere su CLARIN

ILC-CNR – National Executor (nomina MIUR)

Monica Monachini – *CLARIN Italian National Coordinator*

Alessandro Enea – *Data Center*

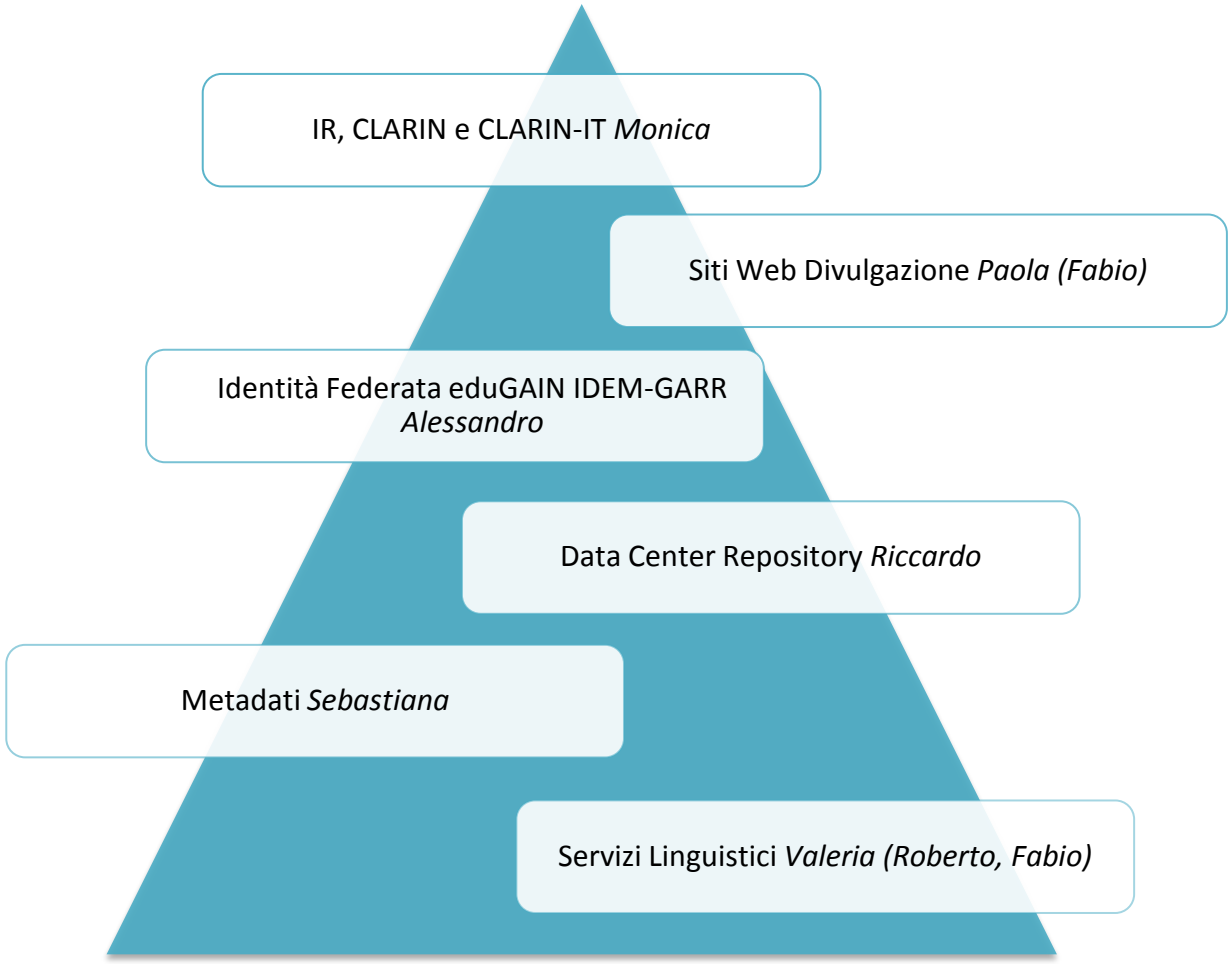
Paola Baroni – *Comunicazione e Divulgazione*

Riccardo Del Gratta – *Repository*

Sebastiana Cucurullo – *Metadati*

Valeria Quochi – *User Involvement*

Schema della presentazione



Sommario



- CLARIN-ERIC: informazioni generali; CLARIN Nazionali: cosa fa un Centro; il Consorzio Nazionale
- Promozione e Comunicazione del Consorzio e del Centro
- Il Centro Nazionale: servizi di accesso, autenticazione; relazioni con IDEM-GARR
- Il repository: servizi di catalogazione, deposito; aspetti tecnici: handle, citazione delle risorse
- La catalogazione: come descrivere una risorsa, metadati, come viene esposta nel VLO
- I servizi linguistici

Infrastrutture: definizioni di partenza



- **rete** di **impianti** e **servizi** interconnessi tra loro da specifici **punti nodali**.
- **rete** di **trasporti**, sia di mezzi, persone, che di materiali, *strade, ferrovie, canali, gasdotti, oleodotti*

… ma anche …



- rete per le **telecomunicazione,** *rete telefonica, televisiva, radiofonica*
- ma anche *la rete informatica*

potenzia le comunicazioni

è un volano per lo sviluppo economico

Le Infrastrutture di Ricerca





- Le Infrastrutture di Ricerca sono **reti di centri**

- forniscono **accesso** internazionale e multidisciplinare a **dati, strumenti e servizi**

favoriscono lo **sviluppo scientifico** e **tecnologico** del Paese

… dal PNR 2015-2020 – MIUR

INFRASTRUTTURE DI RICERCA

Valutare e sostenere in maniera selettiva le infrastrutture di ricerca, pilastro della ricerca internazionale, in particolare della ricerca di base.



- *il PNR investe nella ricerca di base, principalmente attraverso azioni dedicate al capitale umano e alle infrastrutture di ricerca*
- *… obiettivo è quello di dare un sostegno selettivo alle infrastrutture di ricerca.*
- *Il PNR pone grande attenzione alle infrastrutture di ricerca, pilastro fondamentale della ricerca italiana e internazionale, in particolare della ricerca di base.*
- *Il PNR riconosce la necessità di programmare nuove condizioni di contesto per favorire la permanenza dei ricercatori in Italia, a cominciare dagli "ecosistemi" generati dalle Infrastrutture di Ricerca.*

- Le infrastrutture di ricerca (IR) sono tra i **pilastri** della ricerca italiana,
- in particolare della **ricerca di base,** e svolgono un ruolo fondamentale
- **nell' avanzamento della conoscenza**,
- **nello sviluppo dell'innovazione** e delle sue applicazioni, così come
- **nello sviluppo economico** e sociale dei territori nei quali sono insediate.
- … le IR offrono **servizi** qualificati,
- attraggono talenti e
- creano attività di **networking** internazionale,
- contribuendo alla realizzazione di un **ambiente stimolante e competitivo** da cui traggono beneficio, a breve e a lungo termine, le aree che le ospitano.

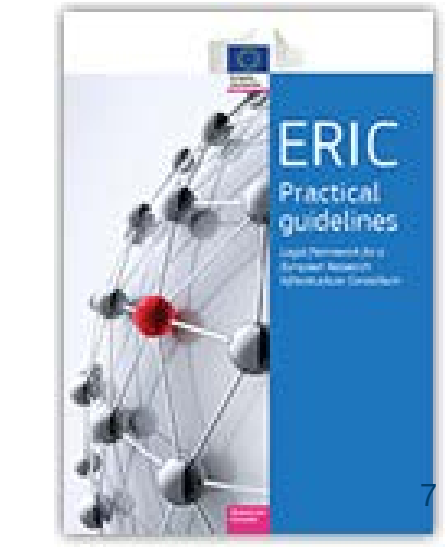
Infrastrutture di Ricerca: organi e strumenti di gestione

L'**ESFRI** (European Strategy Forum on Research Infrastructures) organo creato nel 2002, su volere del Consiglio dell'Unione Europea

- al cui interno delegazioni dei Paesi Membri **definiscono** la **politica** per le IR (fino al 2022).
- ha il compito di realizzare e implementare la **roadmap** strategica per la loro costruzione, operatività e integrazione
- nel 2016 ha identificato anche 29 **landmarks**, infrastrutture mature che rappresentano un riferimento nel settore scientifico tra cui CLARIN

L'**ERIC** (Consorzio Europeo per le Infrastrutture di Ricerca) istituito nel 2009

- rappresenta uno **strumento legale** per facilitare la creazione e il **funzionamento** delle Infrastrutture di Ricerca europee superando gli ostacoli derivanti dai differenti ordinamenti giuridici vigenti nei vari paesi dell'Unione



Cosa si intende per Infrastruttura di Ricerca



FACILITY

Al dato fisico delle strumentazioni, dei servizi e delle funzioni, si aggiunge quello intangibile delle competenze, del saper fare, del facilitare complessivamente un insieme di attività

IR nel settore delle Scienze Umane e Sociali e del Patrimonio Culturale



Infrastrutture Scienze Umane e Sociali e Patrimonio Culturale		
CHARISMA IPERION [MiBaC CNR-DSU]	conservazione e restauro	www.charismaproject.eu
CENDARI [SISMEL]	archivi e risorse per la storia medievale e moderna	www.cendari.eu
DARIAH [MIUR CNR-DSU]	tecnologie digitali per l'arte	www.dariah.eu
ARIADNE [PIN CNR-ISTI/CNR-DSU]	archeologia	www.ariadne-infrastructure.eu
CLARIN [MIUR CNR-DSU]	scienze umane e sociali	www.clarin.eu
EUROPEANA [MiBaC ICCU]	biblioteca digitale europea	www.europeana.eu

CLARIN:
un po’ di storia



- CLARIN nasce da un'idea pionieristica di Zampolli
- La fase preparatoria finanziato come un FP7 nel 2008 prende le mosse da un consorzio di centri di ricerca europei nel settore di ricerca delle "risorse linguistiche"
- CLARIN entra poi a far parte della ESFRI roadmap
- Si costituisce come ERIC nel 2012

1° Ottobre 2015

- L'Italia diventa membro della Infrastruttura CLARIN-ERIC
  Per maggiori informazioni cf. il sito www.clarin.eu
- *Common Language Resource Infrastructure for Social Sciences and Humanities*



- Una opportunità per le Scienze del Linguaggio e le Discipline Umanistiche.

CLARIN-ERIC



- CLARIN ERIC si configura come una federazione di **data centers** di vario tipo, i pilastri dell'infrastruttura dedicata a soddisfare le esigenze della comunità di studiosi delle Scienze Umane e Sociali
- Dal 2012 CLARIN, come gli altri ERIC, non è più finanziato come un progetto europeo, ma diventa autosufficiente ed è sostenuto finanziariamente dai paesi membri
- La nazione che aderisce ogni anno paga all'ERIC una tassa calcolata in base al PIL


Mappa
Satellite
Svezia
Norvegia
Finlandia
Mar Baltico
Mar del Nord
Danimarca
Regno Unito
Irlanda
Polonia
Bielorussia
Mosca
Москва
Germania
Belgio
Ucraina
Moldavia
Ungheria
Francia
Romania
Croazia
Serbia
Italia
Roma
Bulgaria
Mar Nero
Georgia
Portogallo
Madrid
Spagna
Mar Tirreno
Grecia
Istanbul
Turchia
Azerbaiian
Siria
Libano
Tunisia
Mar Mediterraneo
Google

CLARIN-ERIC: i tipi di centro



- **B-Centres, *Service Providing Centres*** offer the scientific community access to resources, services and knowledge on a sustainable basis
- **C-Centres**, ***Metadata Providing Centres***
- K- & L-Centres (general and local ***Knowledge Centres***, part of the CLARIN Knowledge Sharing Infrastructure)
- **T-Centres** (***Trust Centres***, providing access to protected resources via the Service Provider Federation)
- **E-Centres** (***External Centres*** offering central services without being part of any national consortium)

CLARIN-nazionali



Ciascun paese-membro finanzia con fondi propri l'implementazione del CLARIN a livello nazionale.

I CLARIN-nazionali devono:

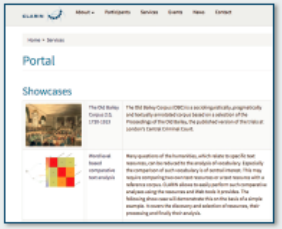
1. impiantare (almeno) un **data center** nazionale di tipo B che offre dati e servizi alla comunità di riferimento → Identifica un Esecutore Nazionale

1. costituire una rete di istituzioni e organizzazioni che costituisce il **consorzio** → Identifica un Coordinatore Nazionale

CLARIN:
i servizi di un data center

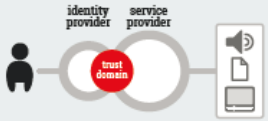
CLARIN Portal

Check out examples
of what is currently feasible



**Easy access
to protected resources**

Use your own institutional credentials



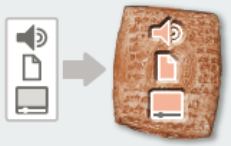
**Language Resource
Inventory**

Browse descriptions
of featured language resources



**Depositing
services**

Store language resources
in sustainable repositories



**Web services
and applications**

Explore and
analyse language data



**Content
Search**

Search within
CLARIN resources
at different locations



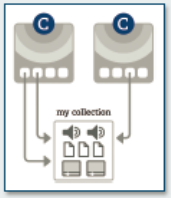
**Virtual Language
Observatory**

Discover language resources



Virtual collections

Combine and
cite data sets



**Consulting
services**

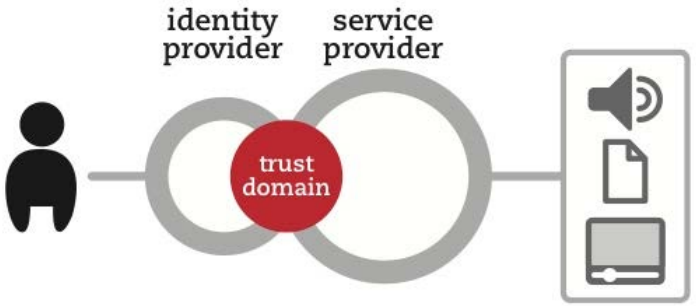
Ask CLARIN about
language resources and tools



CLARIN:
accesso semplice e unificato

Ogni data center deve:

- garantire servizi di **accesso semplice** e **unificato** a risorse protette o disponibili solo tramite autenticazione attraverso il proprio identificativo istituzionale
- garantire attraverso il sistema di autenticazione controllato, la **protezione** delle risorse

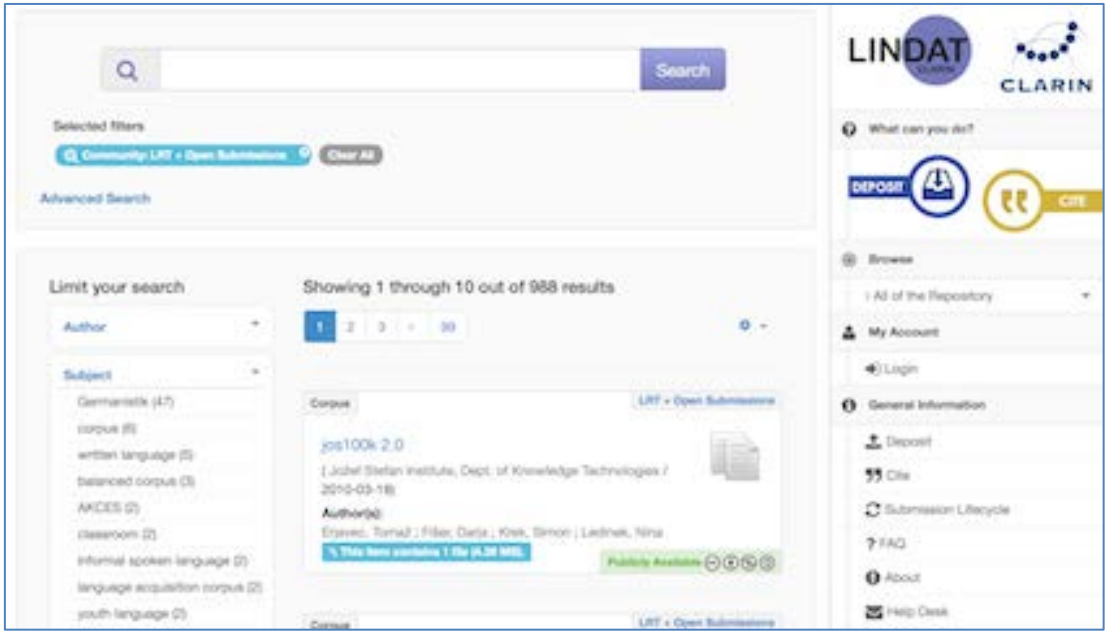




CLARIN:
il repository

Ogni data center deve:

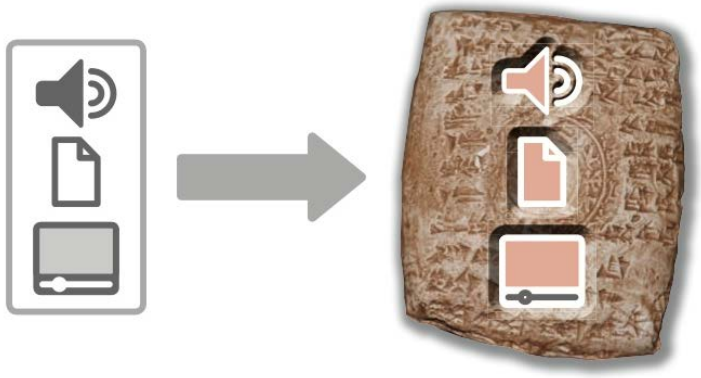
- **creare** un repository
- **depositare** la descrizione tramite metadati
- **documentare** le risorse linguistiche
- **consentire accesso** alle informazioni relative a dati e strumenti di interesse per la ricerca dello studioso
- attribuire un **identificativo persistente** ai dati



CLARIN: deposito

Ogni data center deve:

- garantire servizi di **deposito** di dati e strumenti linguistici in maniera sostenibile
- **archiviazione** a lungo termine
- **descrizione** dei dati in maniera univoca e ricercabile attraverso l'infrastruttura



CLARIN: il VLO

Ogni data center deve:

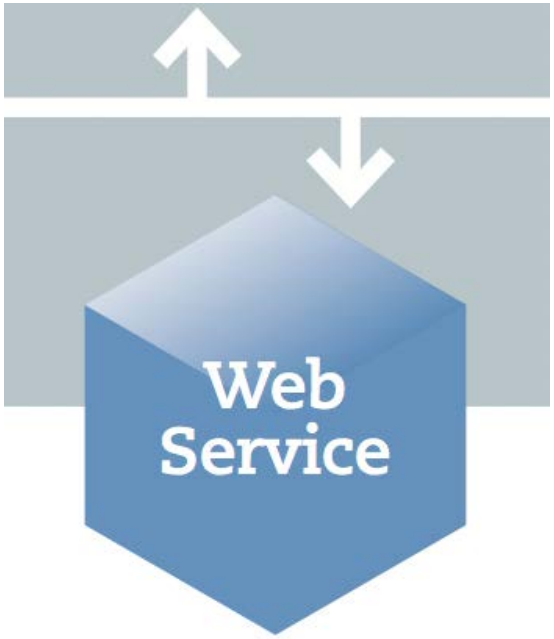
- Garantire servizi di **harvesting** che consentano di riversare i (meta)dati dai repository locali nel catalogo centrale
- **scoprire** le risorse linguistiche attraverso una ricerca libera oppure facilitata tramite il sistema a faccette



CLARIN: applicazioni NLP

Ogni data center deve:

- fornire una varietà di strumenti che consentano di **esplorare** i dati linguistici incapsulati sotto-forma di **servizi** per i quali non c'è necessità di competenze tecniche
- integrare i propri servizi all'interno di ambienti di annotazione promossi in CLARIN come per esempio WebLicht fornito dal centro tedesco di *Tübingen*



CLARIN-ERIC: organi gestionali

https://www.clarin.eu/content/governance



General Assembly: Massimo organismo decisionale. Responsabili designati dal ministero di ogni nazione.

Clarin office + **Board of directors**: gestione esecutiva



National coordinators' forum: coordinatori dei consorzi nazionali

+

Standing committee for technical centers: responsabili tecnici dei data-center nazionali



CLARIN-IT:
chi fa cosa (nomina MIUR)



- Il Dipartimento di Scienze Umane e Sociali, Patrimonio Culturale del CNR (DSU-CNR)
  - è l'organo di rappresentanza
  - rappresenta l'Italia alla Assemblea Generale nella persona del Prof. Riccardo Pozzo
- Il Coordinatore Nazionale, Monica Monachini (ILC-CNR)
  - ha il compito di creare il Consorzio Nazionale,
  - fungere da interfaccia con l'ERIC, partecipare/nominare esperti nei vari comitati
  - identificare esperti nazionali per la partecipazione a eventi tematici organizzati dall'ERIC
- L'Istituto di Linguistica Computazionale "A. Zampolli" del CNR (ILC-CNR)
  - è l'istituto esecutore
  - ha il compito di implementare il data center con un repository di dati e strumenti linguistici e fornire servizi alla comunità di riferimento

ILC4CLARIN:
il Data Center presso l'ILC



- Coordinamento Tecnico: partecipare allo standing committee; gestire le attività tecnologiche del centro
- Amministrazione del repo: mantenere contatti con esperti di CLARIN-Dspace, partecipare ai WS relativi al software, seguire harvesting MD, mantenere il repo
- Gestione comunicazione e disseminazione: curare il sito web del data center
- Gestione MD: curare procedure inserimento e validazione
- Implementazione servizi: curare l'adattamento e integrazione dei servizi esistenti, sviluppare nuovi servizi

CLARIN-IT: il Consorzio



- Il consorzio nazionale, resp. Coordinatore Nazionale (NC), è costituito da università, istituti di ricerca, biblioteche e archivi
- i partners del consorzio mettono a disposizione della comunità di riferimento dati e servizi
- il consorzio, tramite il NC, assicura l'implementazione delle strategie indicate annualmente dalla Assemblea Generale, all'interno della propria nazione
- esperti nazionali, di comune accordo con il NC, partecipano ai vari Comitati Tecnici dell'ERIC
- lo stato delle attività del consorzio sono riportate annualmente all'Assemblea generale dal NC

CLARIN-IT: primo nucleo



CLARIN-IT		
Università di Siena	archivi orali	Silvia Calamai
Scuola Normale Superiore	archivi orali	Pier Marco Bertinetto
Università di Siena	archivio della latinità del medioevo	Francesco Stella
EURAC Bolzano	dati e strumenti per le lingue regionali	Andrea Abel
FBK Trento	strumenti per applicazioni di NLP	Bernardo Magnini e Sara Tonelli
Università Cattolica Milano	strumenti per le lingue classiche	Marco Passarotti
Università Ca' Foscari	annotazione tematica in DH	Paolo Mastandrea
Università di Parma	edizioni digitali per il greco	Anika Nicolosi
Università di Pisa	dati e strumenti per NLP	Alessandro Lenci
Università di Roma	ontologie per DH	Fabio Ciotti/D. Silvi

CLARIN-IT:
come 'ritornano' le fees



CLARIN-ERIC, tramite CLARIN-PLUS, il braccio operativo, all'interno dei vari Consorzi:

- favorisce lo scambio di competenze tra studiosi
- offre brevi soggiorni di ricerca presso centri specializzati tramite lo strumento dei Mobility Grant
- favorisce la partecipazione dei partners a giornate di lavoro su temi strategici condivisi dalla Agenda Strategica di CLARIN
- promuove l'organizzazione e/o la partecipazione a eventi
  - CLARIN-PLUS Workshop on Spoken Word Data in Oral History
    (Oxford, aprile 2016): UniSi + ILC
  - CLARIN-PLUS Workshop on Digital Collections of Newspapers
    (Leuven, settembre 2016): FBK
  - CLARIN Workshop on Multilingual Corpus Annotation Tools
    (Ljubljana, novembre 2016): ILC (+ FBK)
  - CLARIN Workshop on Interoperability of Lexico-Semantic Resources
    (Tartu, gennaio-febbraio 2017): ILC
  - CLARIN-PLUS Workshop on Parliamentary Records
    (Sofia, 27-29/03/2017): Senato + ILC

CLARIN-IT:
il sito Web del Consorzio



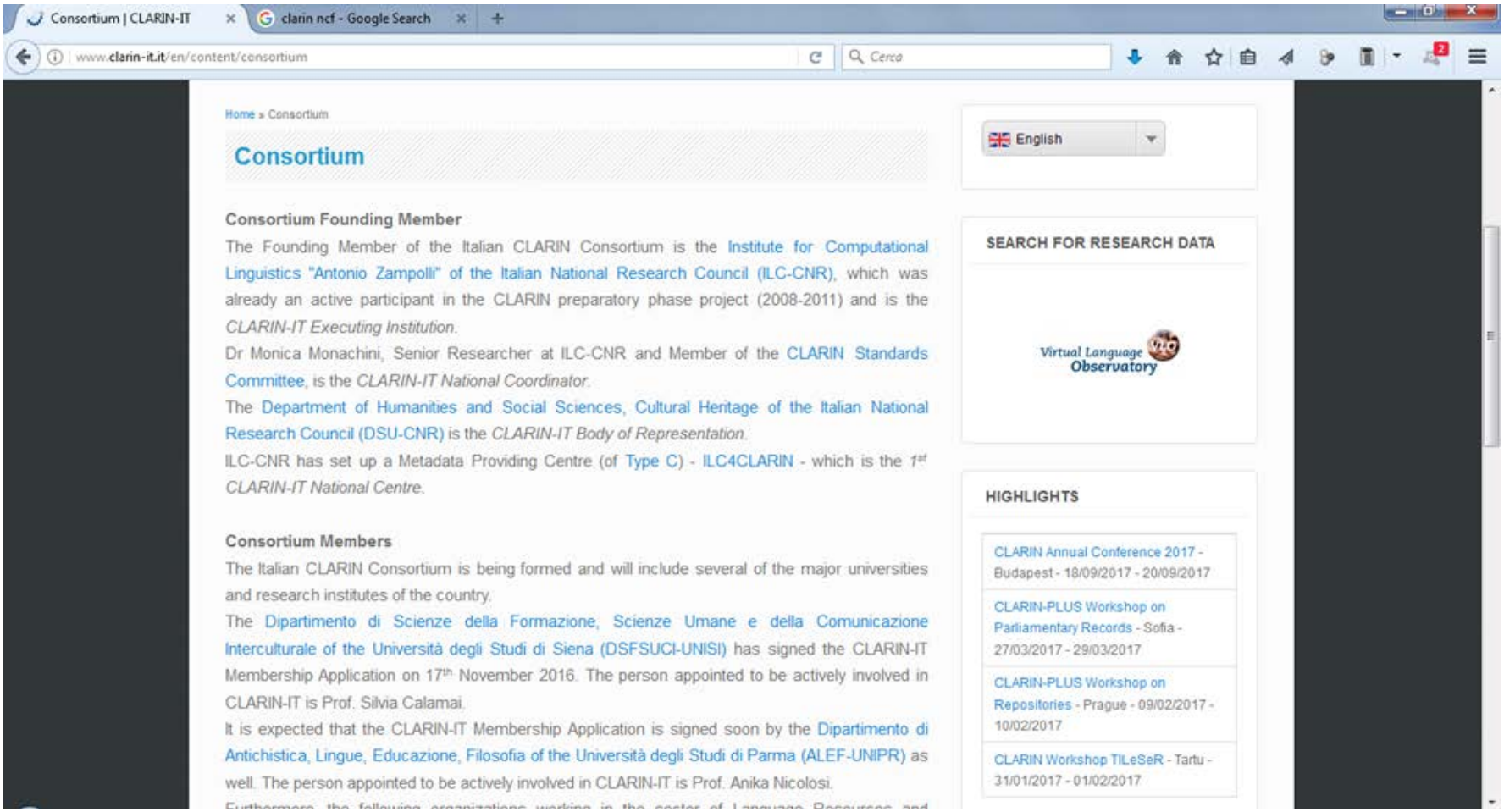
- http://www.clarin-it.it
  - è la **'vetrina'** del Consorzio Italiano
    - cos'è e com'è nato
    - com'è strutturato e come funziona (**governance** 'ispirata' a quella di CLARIN, in progress)
    - da quali **membri** è costituito (in progress)
    - quali **vantaggi** offre alla comunità delle Scienze Umane e Sociali
  - contiene **informazioni e materiale relativi agli eventi** organizzati da CLARIN
    - incontri degli organi e dei comitati di CLARIN (General Assembly, National Coordinators' Forum, Standing Committee for CLARIN Technical Centres ecc.)
    - conferenze annuali di CLARIN e workshop organizzati da CLARIN, CLARIN-PLUS e Centri CLARIN

CLARIN-IT:
il sito Web del Consorzio

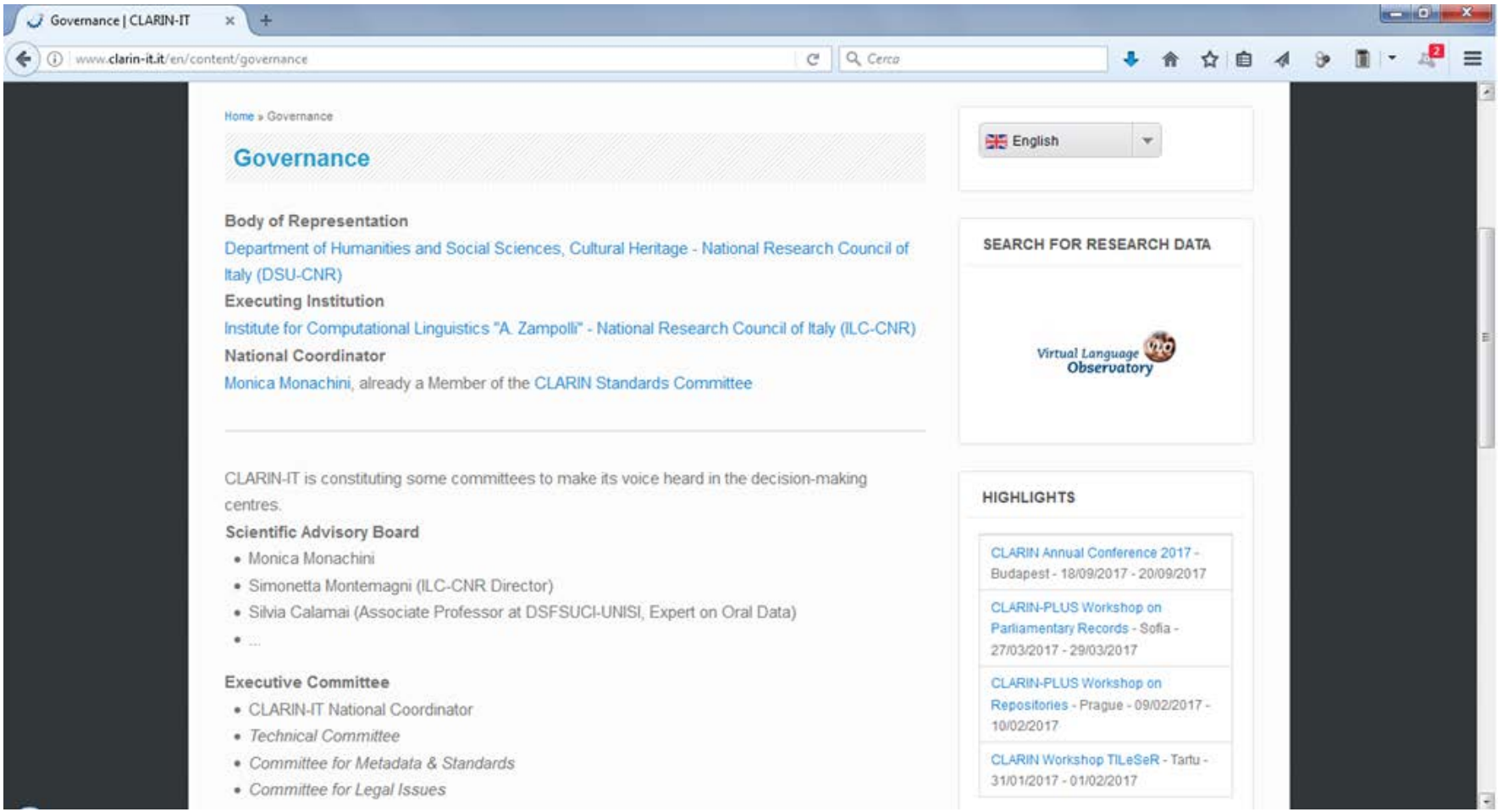


- http://www.clarin-it.it
  - promuove la **partecipazione agli eventi** organizzati da CLARIN
    - call for papers per le conferenze annuali di CLARIN
    - call for proposals per workshop organizzati da CLARIN e CLARIN-PLUS
    - call for papers per workshop organizzati dai Centri CLARIN
  - inoltre
    - contiene notizie relative alla partecipazione di CLARIN-IT agli eventi organizzati da CLARIN
    - promuove la partecipazione agli eventi più importanti del settore delle Scienze Umane e Sociali
    - consente l'accesso diretto al Virtual Language Observatory (VLO) di CLARIN
    - contiene un collegamento diretto al sito del Centro ILC4CLARIN

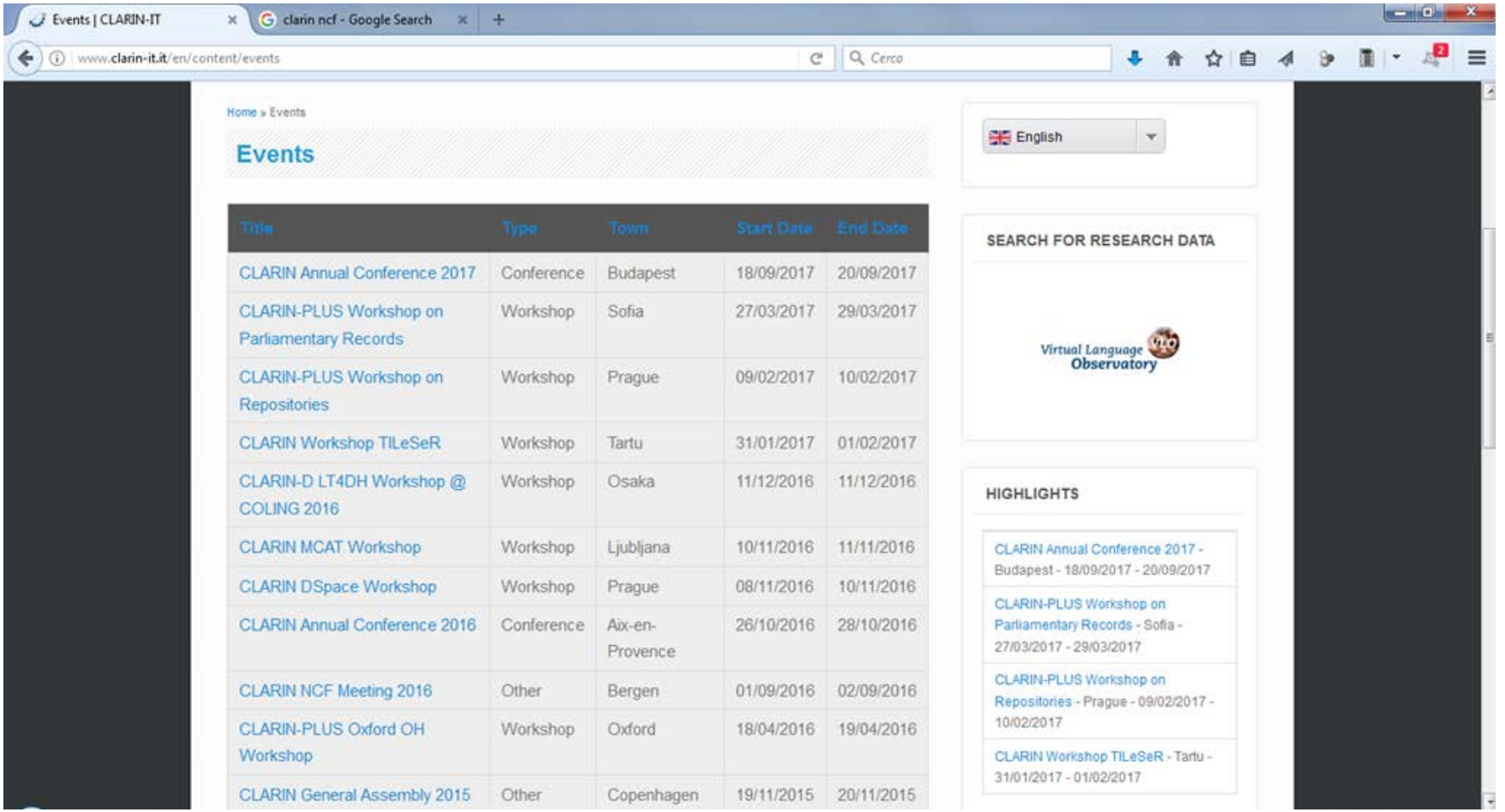
Il Consorzio:
Membri (in progress)



La Governance del Consorzio: Organi e Comitati (in progress)



Informazioni e Materiale relativo agli Eventi organizzati da CLARIN



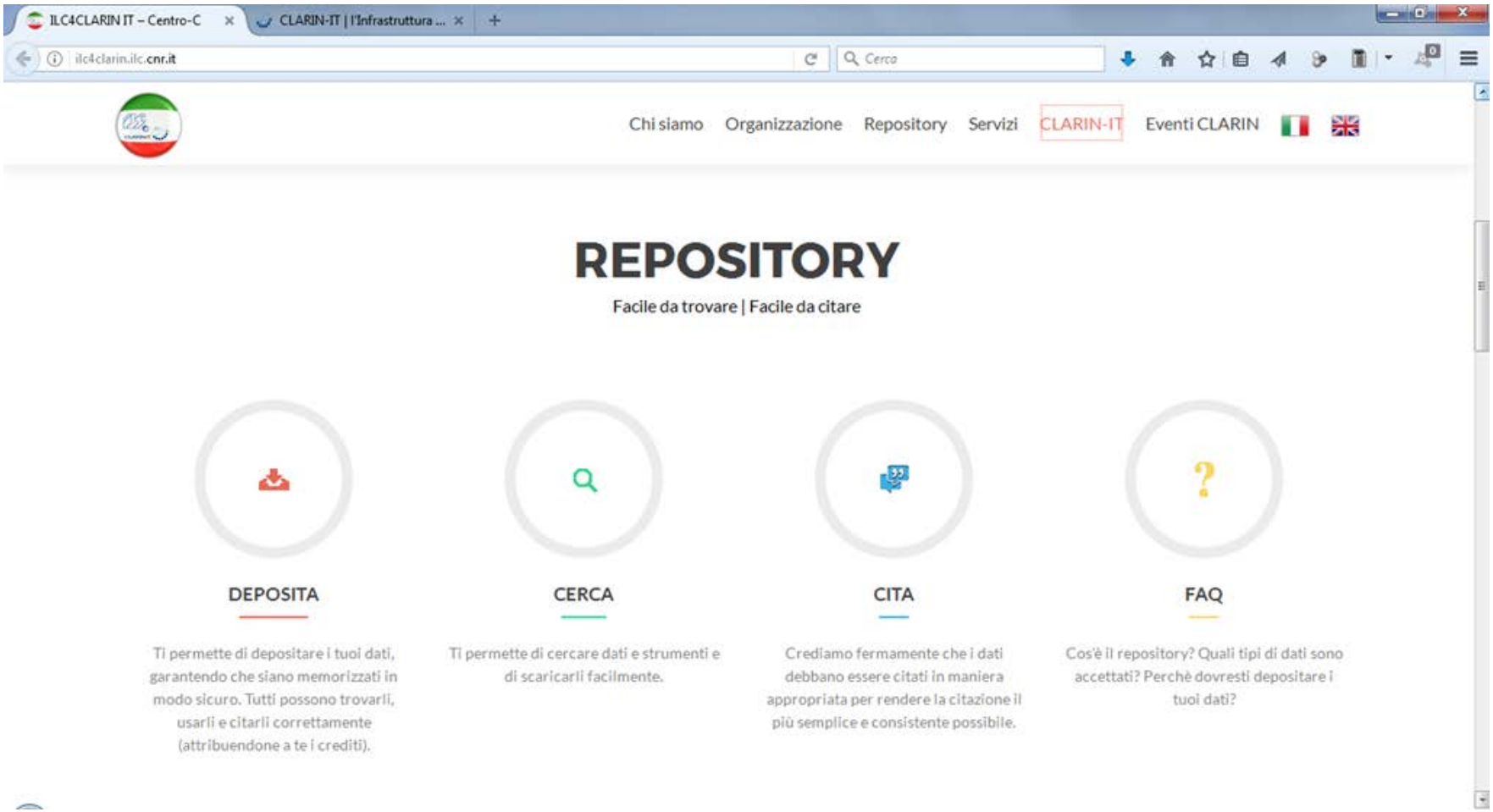
Title	Type	Town	Start Date	End Date
CLARIN Annual Conference 2017	Conference	Budapest	18/09/2017	20/09/2017
CLARIN-PLUS Workshop on Parliamentary Records	Workshop	Sofia	27/03/2017	29/03/2017
CLARIN-PLUS Workshop on Repositories	Workshop	Prague	09/02/2017	10/02/2017
CLARIN Workshop TILeSeR	Workshop	Tartu	31/01/2017	01/02/2017
CLARIN-D LT4DH Workshop @ COLING 2016	Workshop	Osaka	11/12/2016	11/12/2016
CLARIN MCAT Workshop	Workshop	Ljubljana	10/11/2016	11/11/2016
CLARIN DSpace Workshop	Workshop	Prague	08/11/2016	10/11/2016
CLARIN Annual Conference 2016	Conference	Aix-en-Provence	26/10/2016	28/10/2016
CLARIN NCF Meeting 2016	Other	Bergen	01/09/2016	02/09/2016
CLARIN-PLUS Oxford OH Workshop	Workshop	Oxford	18/04/2016	19/04/2016
CLARIN General Assembly 2015	Other	Copenhagen	19/11/2015	20/11/2015

ILC4CLARIN:
il sito Web del Centro

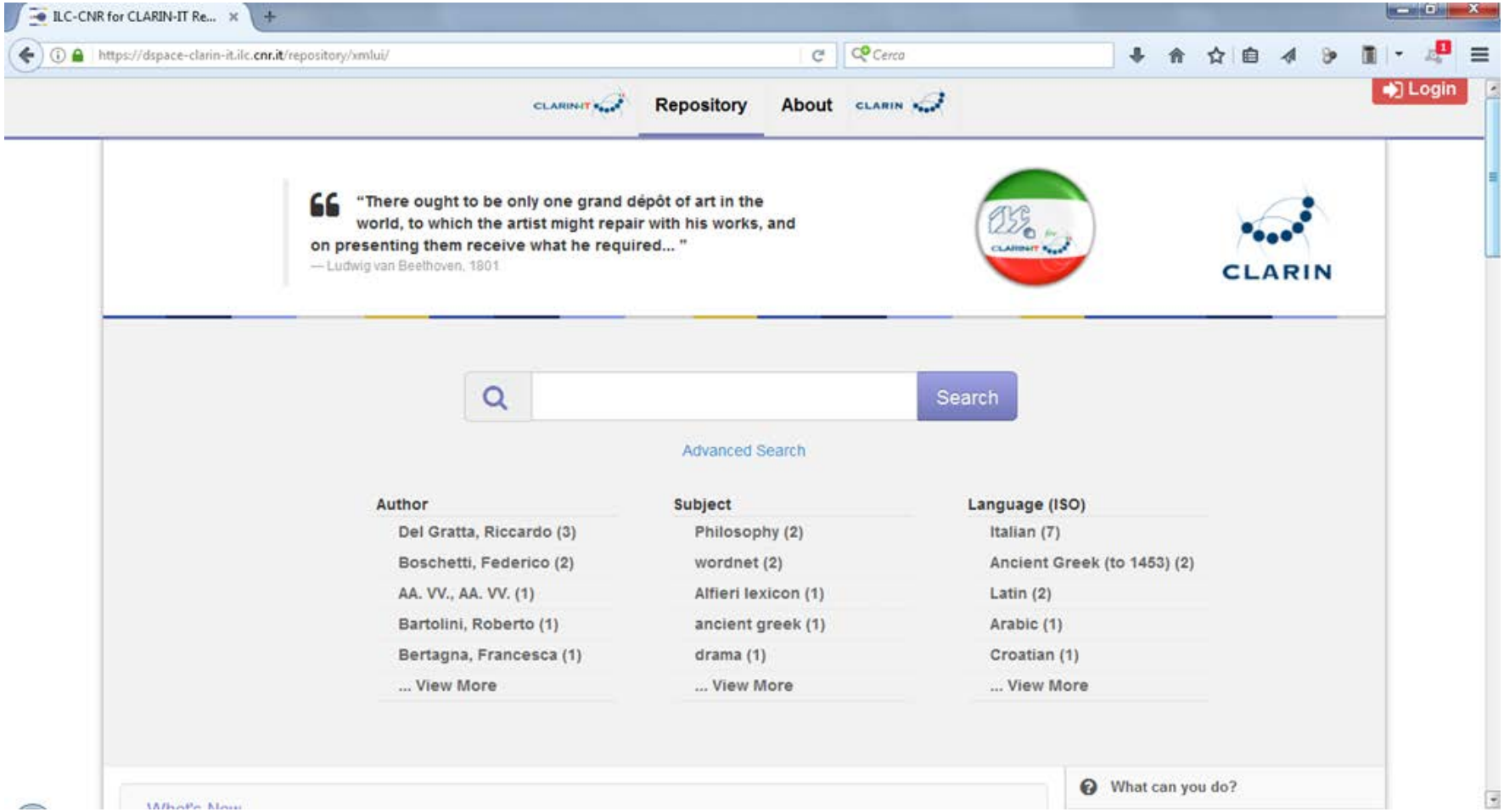


- https://ilc4clarin.ilc.cnr.it
  - è il sito del **1° Data Center** del Consorzio Italiano
  - contiene collegamenti semplificati
    - alle **funzioni del repository** del Centro (deposito, ricerca, citazione)
    - alle FAQ
  - contiene collegamenti diretti
    - al **repository** del Centro
    - alla **griglia dei servizi** del Centro (in progress)
    - al sito del Consorzio Italiano
    - agli eventi organizzati da CLARIN

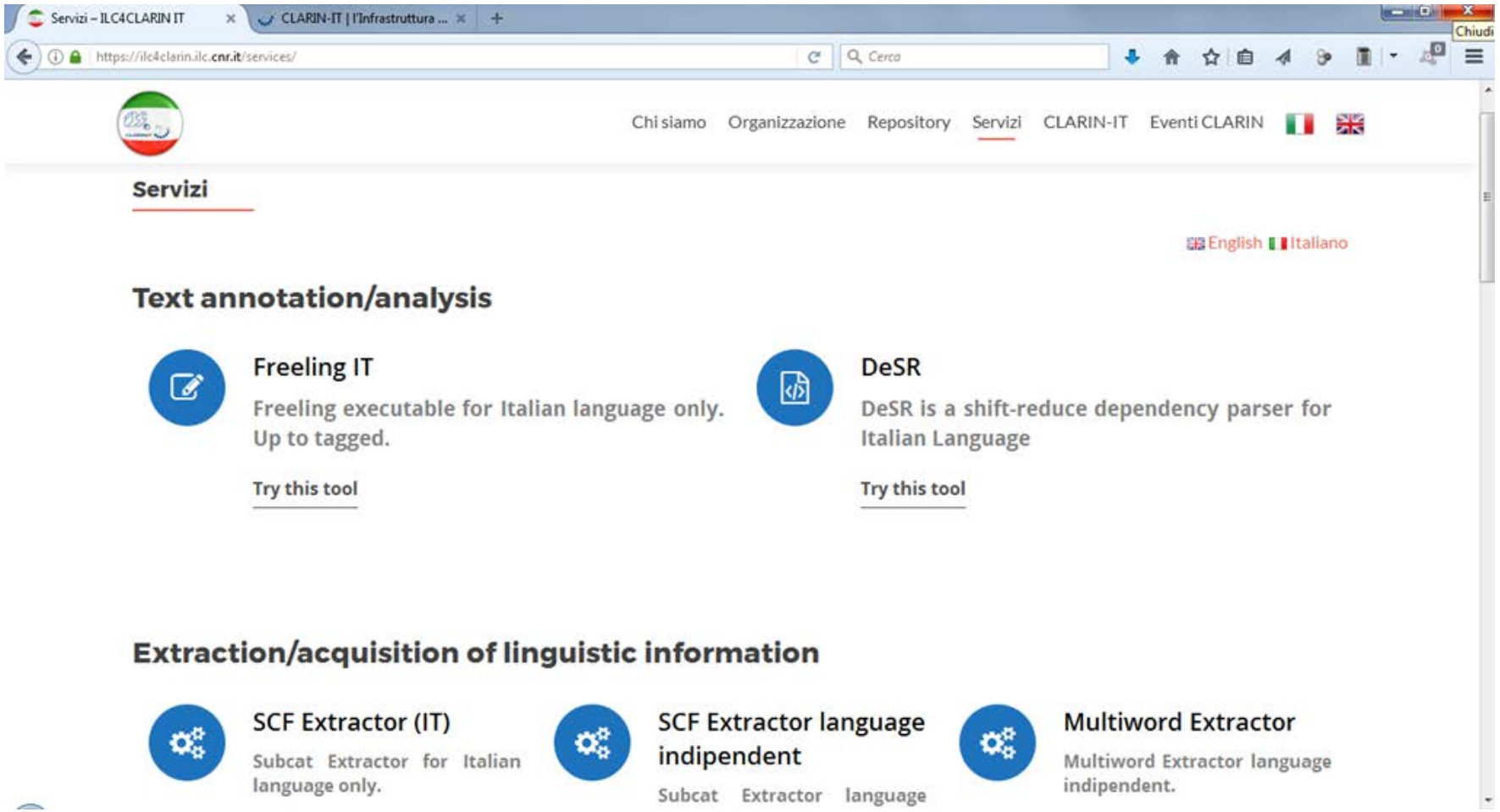
Collegamento Semplificato alle Funzioni del Repository



Collegamento Diretto al Repository



Collegamento Diretto alla Griglia dei Servizi (in progress)



Identità Federata: Single sign-on, definizioni



Dalla pagina: https://www.clarin.eu/content/clarin-in-a-nutshell

It is a research infrastructure that was initiated from the vision that all digital language resources and tools from all over Europe and beyond are accessible through a single sign-on online environment for the support of researchers in the humanities and social sciences.

Single sign-on :Access to multiple systems by logging in once

AAI: Authentication and Authorization Infrastructure

Dettagli su: https://www.clarin.eu/content/federated-identity

The user stores from universities and academic institutions ("Identity Providers") are connected to password-protected web applications ("Service Providers").

Dettagli su: https://www.clarin.eu/content/service-provider-federation

Accesso alle risorse protette con l'Identità Federata: https://www.clarin.eu/content/easy-access-protected-resources

Identità Federata: SSO con IDEM-GARR



In Europa eduGAIN, in Italia IDEM-GARR

- In Italia IDEM-GARR: https://www.idem.garr.it
- ILC-CNR è in IDEM-GARR dalla sua nascita
- Risorse della Federazione IDEM: https://www.idem.garr.it/servizi/sp
- Gestione e Verifica delle Identità in IDEM: https://www.idem.garr.it/servizi/idp

http://www.garr.it

Per accedere: <*nome.cognome*> + password posta ILC

Il centro nazionale



- Centri CLARIN: https://www.clarin.eu/content/clarin-centres
- C-Centres (***Metadata Providing Centres***, their metadata are integrated with CLARIN but they need not to offer any further services)
- https://www.clarin.eu/content/overview-clarin-centres

- ILC4CLARIN Repository: https://www.clarin.eu/content/repositories
- Accesso Federato: https://dspace-clarin-it.ilc.cnr.it

Cosa è un repository



- Un repository è un luogo centralizzato dove i dati (eventualmente in forma aggregata) sono catalogati, salvati e mantenuti.
- In ILC4CLARIN usiamo **clarin-dspace**, una versione personalizzata di **dspace** sviluppato nel data center della repubblica Ceca.
- Attraverso **clarin-dspace** si catalogano risorse e strumenti linguistici, servizi e applicazioni web.

Cosa è un repository



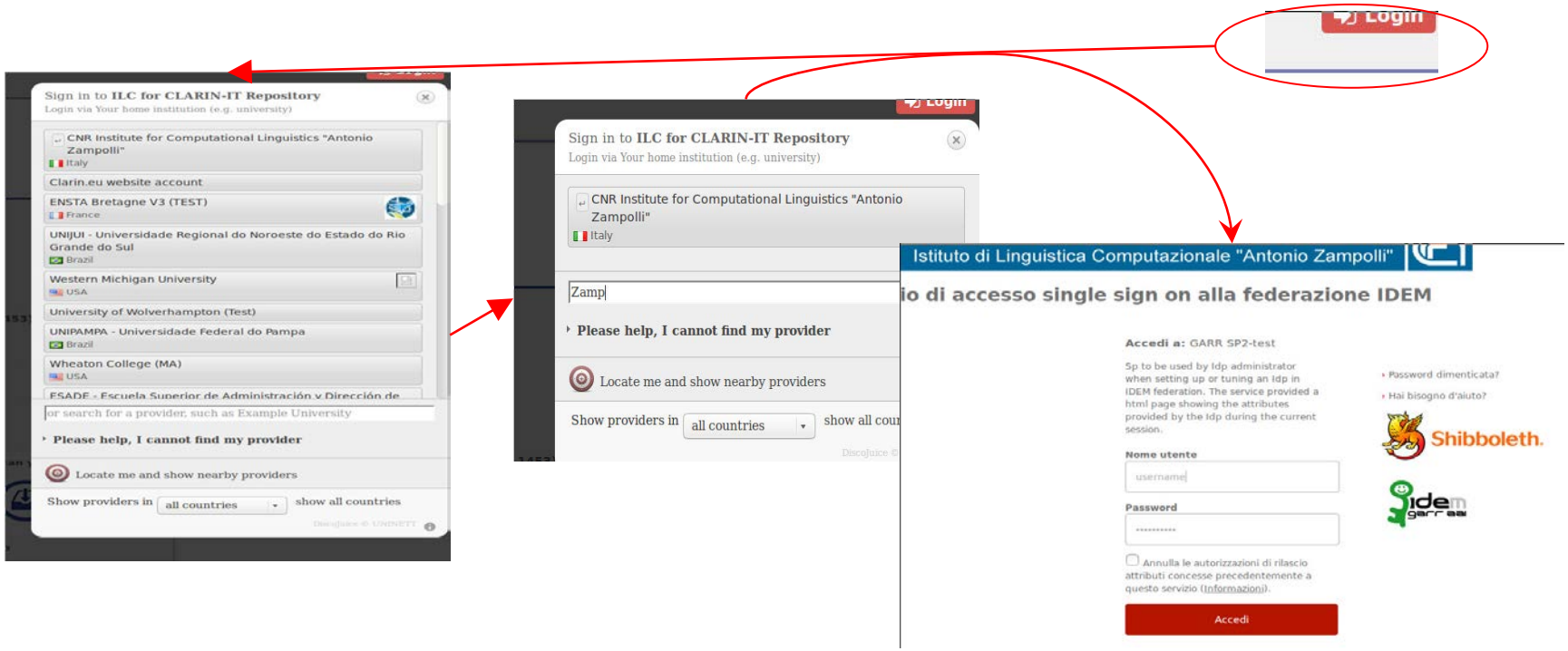
- Il repository è diviso in comunità e collezioni. Comunità e Collezioni sono usati dal repository per creare una struttura gerarchica nel repository:
  - Comunità
    - Collezione 1
    - Collezione 2
    - ....

Le collezioni sono importanti per i seguenti motivi:

- Gli utenti vengono accreditati ad effettuare operazioni sulle collezioni;
- Lo workflow di inserimento e validazione dei metadati sono decisi a livello di collezione.
- Due comunità: ILC per le risorse interne; OPEN per gli altri membri del consorzio
- ILC ha collezione ILC data & tools

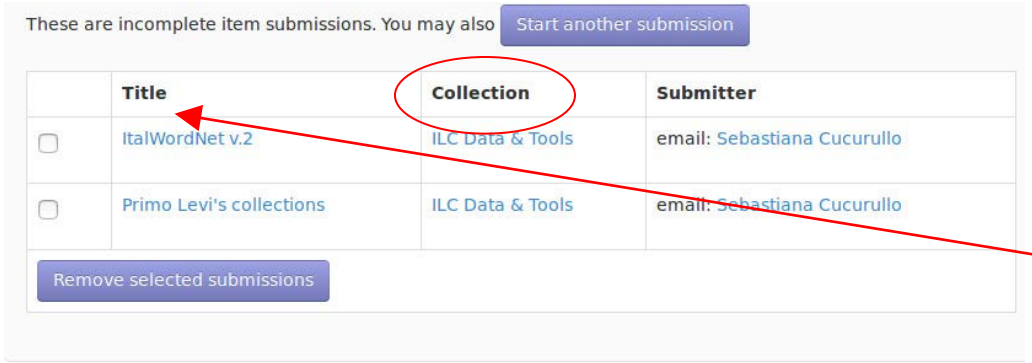
Flusso di catalogazione: login

- Home del repository: **https://dspace-clarin-it.ilc.cnr.it/**
- Per descrivere le risorse all'interno del repository, occorre autenticarsi.
- Dopo login si può accedere alla propria pagina personale: **https://dspace-clarin-it.ilc.cnr.it/repository/xmlui/mydspace**



Flusso di catalogazione: pagina personale

- La pagina personale raccoglie la lista di attività iniziate e/o terminate dall'utente



Sottomissioni non complete

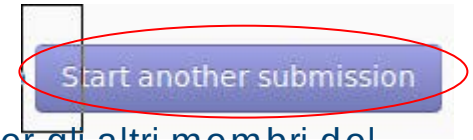


Sottomissioni terminate ed archiviate nel repository

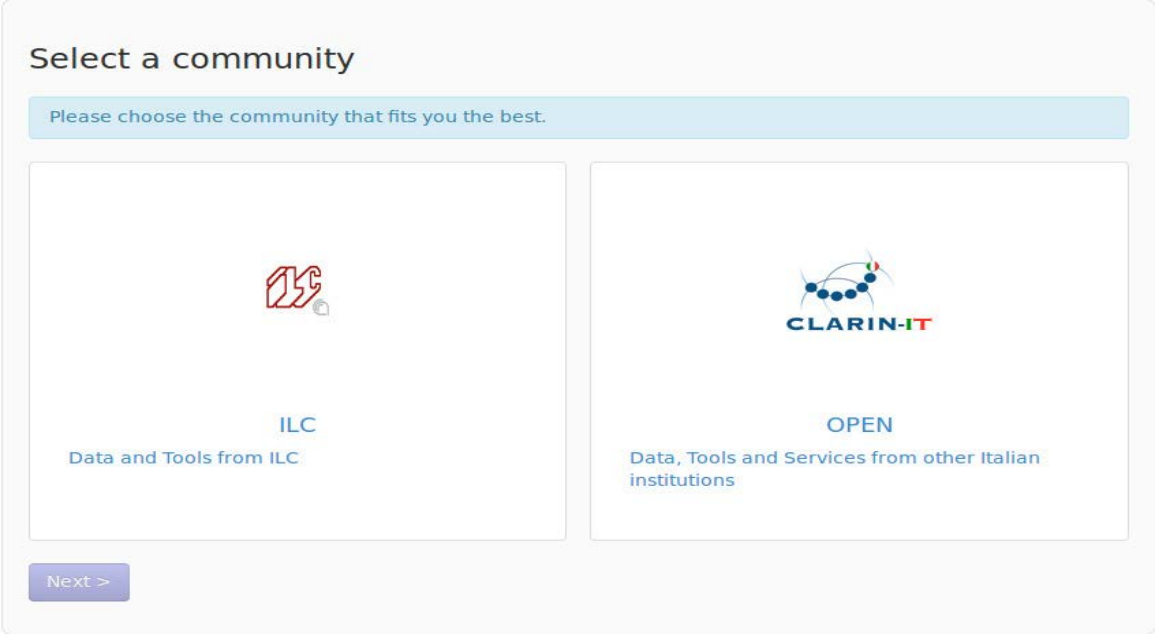
Flusso di catalogazione: iniziare selezionando ‘comunità’ e ‘collezione’

Dopo aver fornito le credenziali per il sistema di SSO

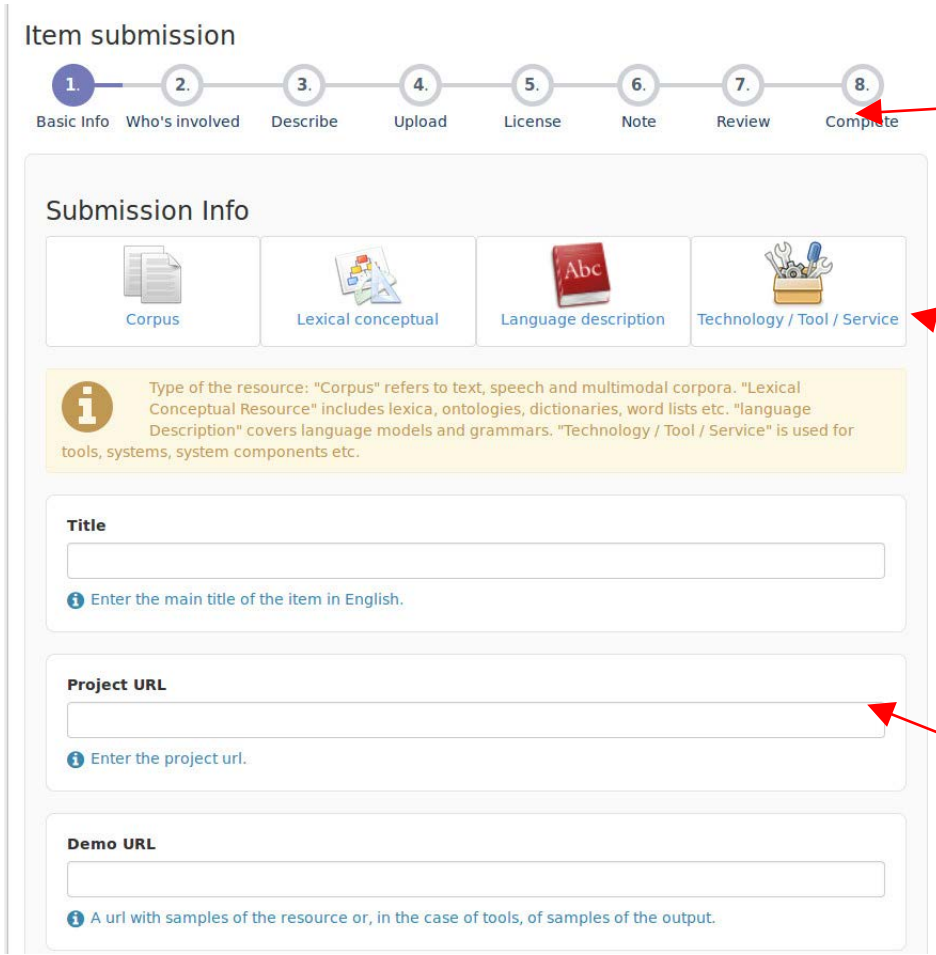
- cliccare su submission e poi “Start another submission”



- Selezionare la comunità (ILC per le risorse interne; OPEN per gli altri membri del consorzio)



Flusso di catalogazione: passo 1) tipo di risorsa

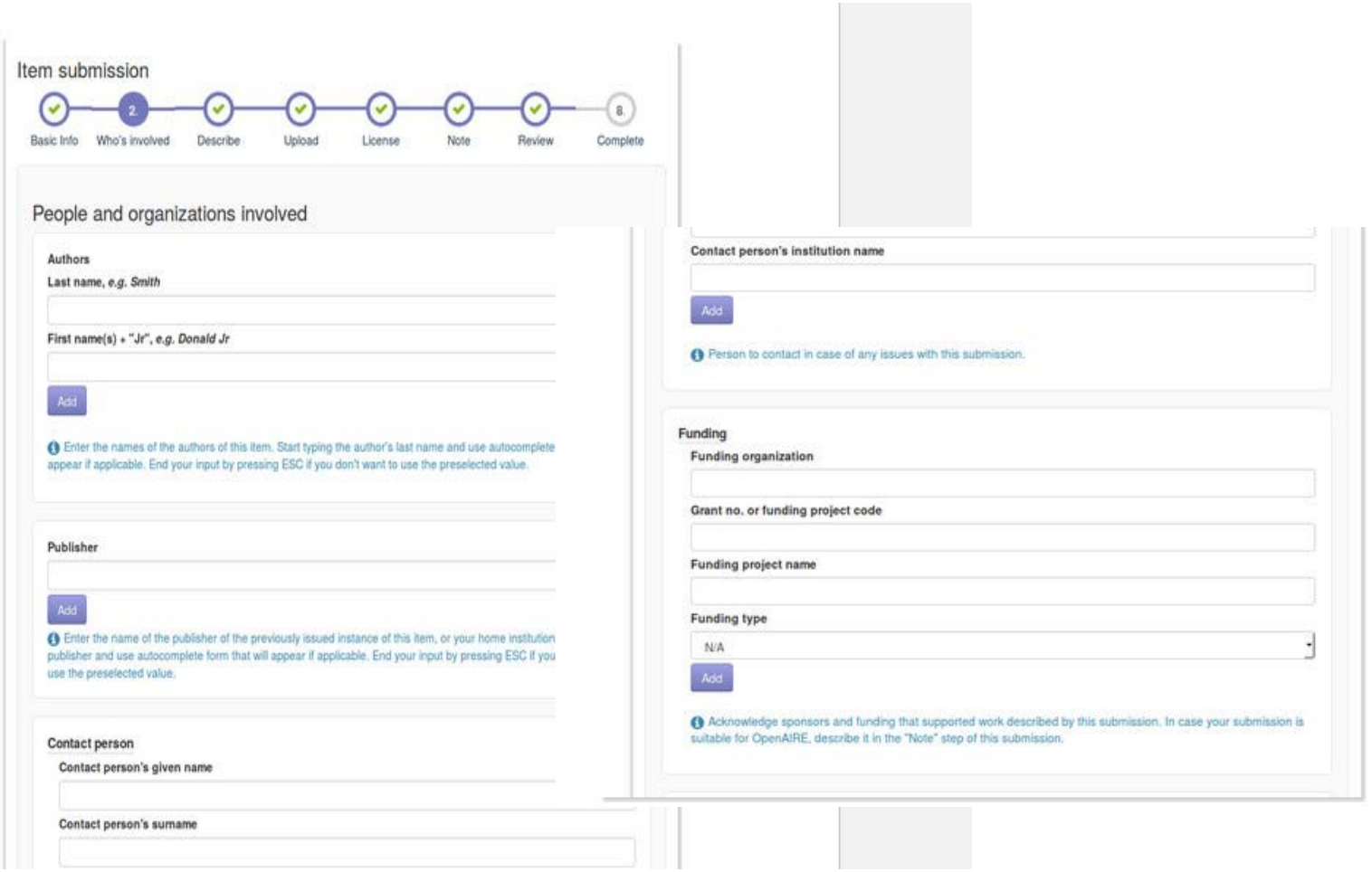


Workflow che guida l'utente nella catalogazione

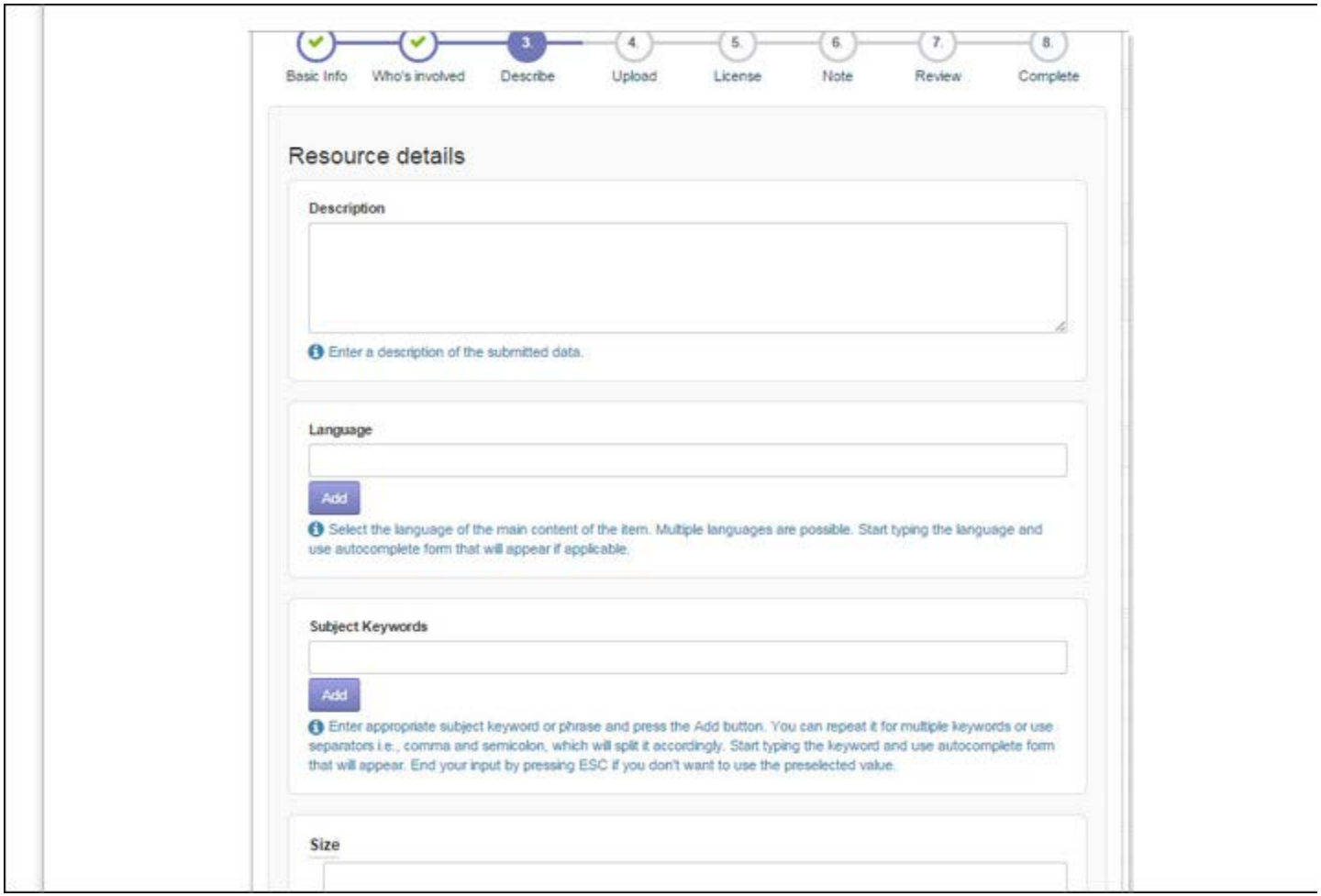
Tipi di risorse

Metadati descrittivi

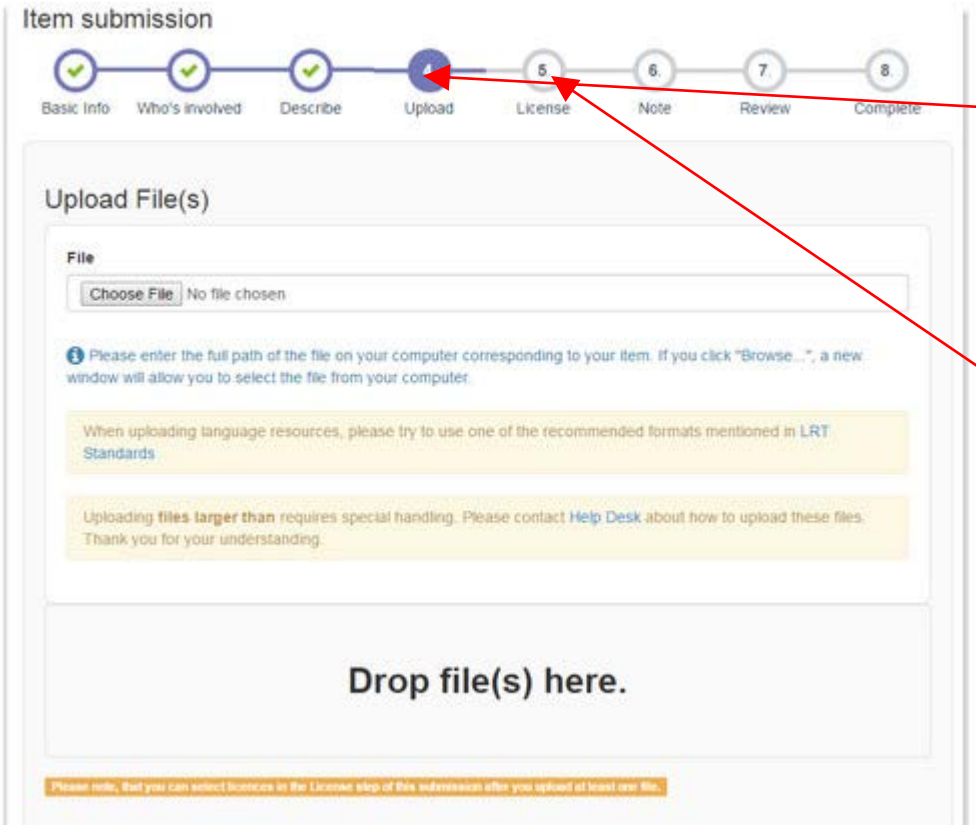
Flusso di catalogazione: passo 2) chi è coinvolto (istituzioni/autori, …)



Flusso di catalogazione: passo 3) dettagli della risorsa (lingua, parole chiave, ...)



Flusso di catalogazione: passo 4) deposito di uno o più file



Associare file alla scheda determina un servizio di deposito: i file vengono salvati sul file system del repository

Se si depositano file è obbligo depositare una licenza [5]

Flusso di catalogazione: passo 5.1) licenza di distribuzione

- Accettare la licenza di distribuzione dei metadati è obbligatorio

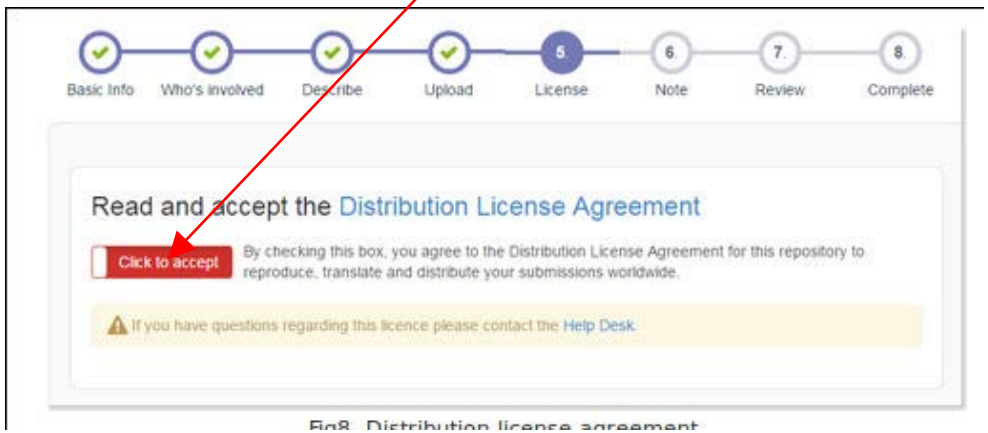
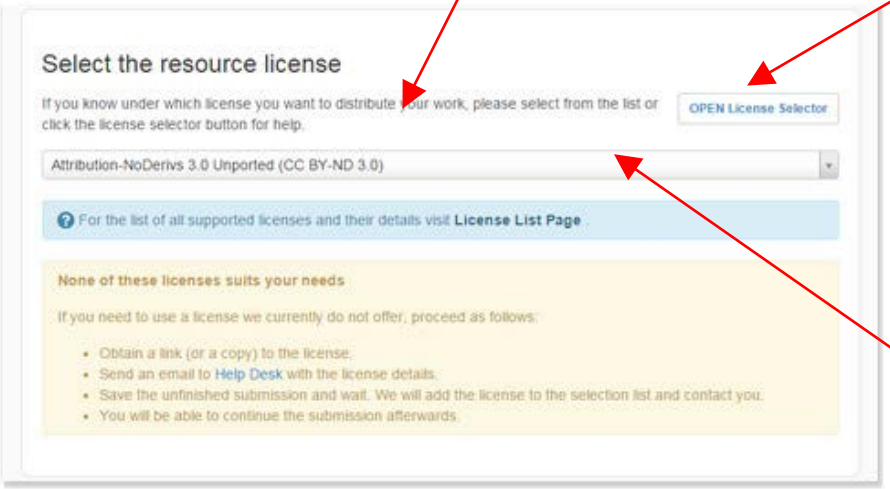


Fig8. Distribution license agreement

Flusso di catalogazione: passo 5.2) selezionare una licenza

- Nel caso si associno file si deve selezionare una licenza per file

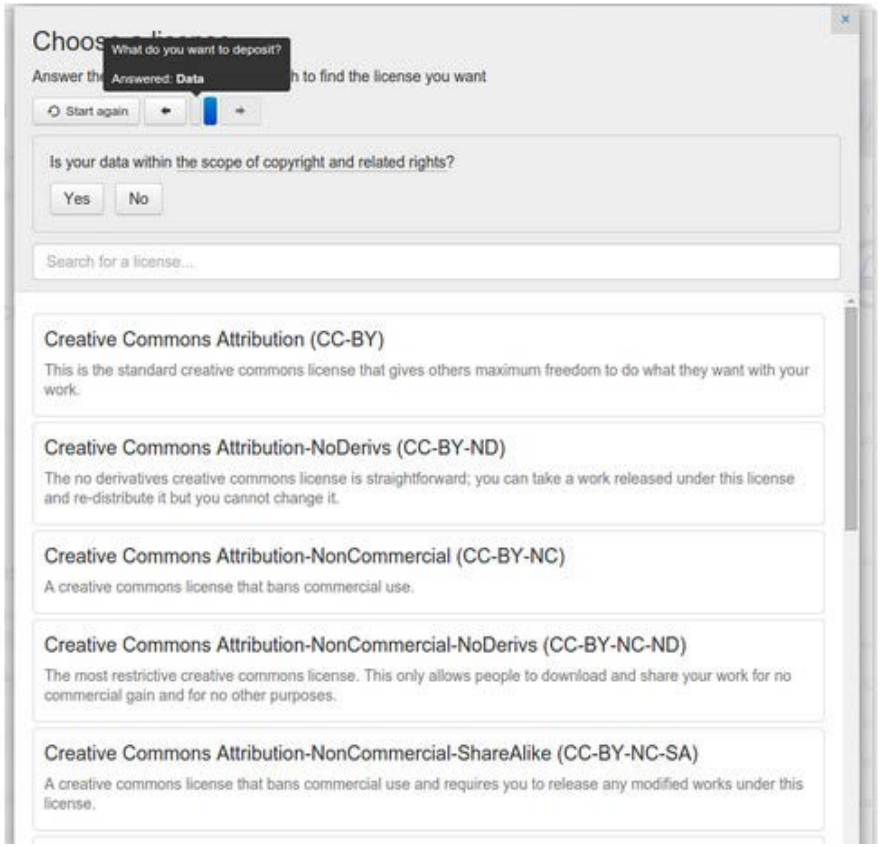


Il selettore permette all'utente di cercare una licenza in base a delle caratteristiche specifiche della stessa (vedi dopo)

Una licenza aggiuntiva è necessaria nel caso si depositino dei file

Flusso di catalogazione: passo 5.3) selettore di licenze

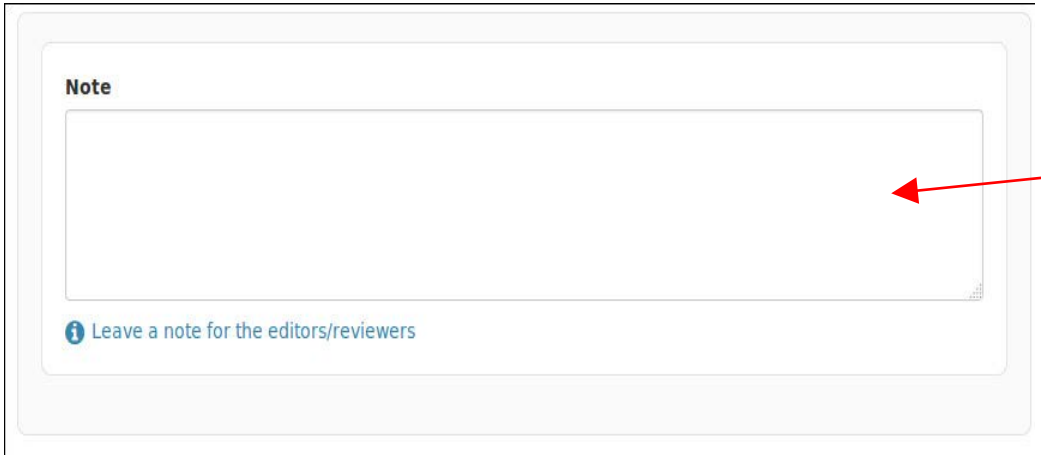
Attraverso una serie di domande il sistema propone delle licenze



IMP: E' comunque sempre possibile definire la propria licenza d'uso per specifiche risorse

Flusso di catalogazione: passi 6) 7) 8) note addizionali

I passi 6, 7 e 8 permettono all'utente di aggiungere delle note interne, rivedere i propri dati e correggerli e finalizzare la sottomissione



Note per revisori/editori

Link utili:

- https://dspace-clarin-it.ilc.cnr.it/repository/xmlui/page/deposit
- https://dspace-clarin-it.ilc.cnr.it/repository/xmlui/page/faq

Flusso di catalogazione: bottoni utili

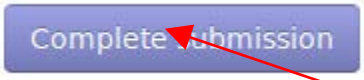
Ogni passo ha una serie di bottoni che permettono all'utente che inserisce di interagire col sistema e con altri utenti



Il bottone next permette di passare alla schermata successiva. Il sistema controlla se i campi obbligatori sono stati riempiti e se campi particolari come email e URL sono formalmente corretti

Il bottone Save&Exit salva la sottomissione come temporanea nella **home** dell'utente. Questa può essere ripresa in seguito attraverso una funzione di **resume**

Il bottone Save&Share permette di **passare** la propria sottomissione ad un altro utente. Il sistema genera un link che viene spedito via mail all'utente e da questo ad un'altra persona



Il bottone Complete assegna la sottomissione agli utenti interni del repository: esperti di metadati e amministratori

Workflow di validazione metadati

- Definito a livello di collezione
- Coinvolge diversi tipologie di validatori:
  - Esperto di metadati;
  - amministratore del repository



Gli esperti di metadati controllano la scheda ed eventualmente la rigettano nel caso questa contenga dei valori errati.

Se la scheda è corretta, gli esperti di metadati la consegnano all'amministratore del repository



Gli amministratori del repository eseguono il "commit" nel repository e questo determina le seguenti cose:

- Creazione dell'Handle
- Inserimento della nuova scheda nel sistema di harvesting (OAI-PHM)

Flusso di catalogazione: parte amministrativa

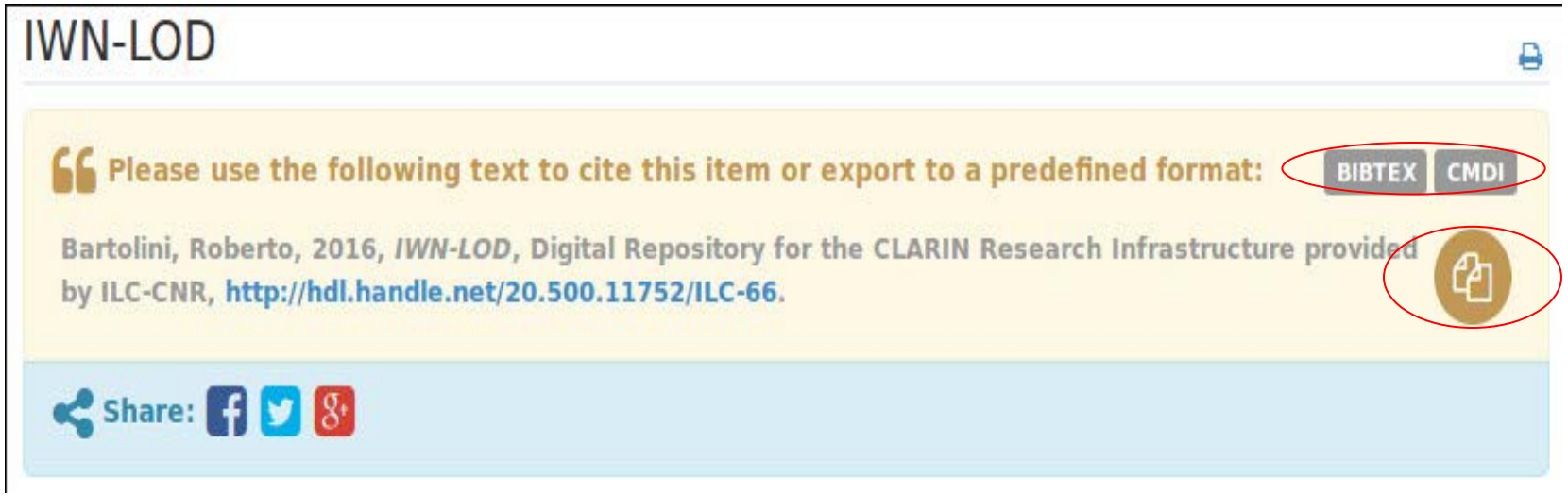


- Cosa otteniamo dopo l'avvenuta registrazione di una scheda nel repository?
  - –handle (persistenza, versionamento, replicabilità)
  - –citazione delle risorse

- Cosa è un identificativo persistente? (https://www.clarin.eu/content/persistent-identifiers)
  - –Un identificativo "associato" ad una risorsa, sia essa una scheda di metadati in un repository, una pagina web o un file fisico che garantisce
    - •la "disponibilità" della risorsa in rete;
    - •la citabilità della risorsa in modo standard.

Handle System

- Un PID garantisce
  - **persistenza:** il pid garantisce che l'URI della risorsa venga sempre risolto, aggirando il problema connesso allo spostamento fisico della risorsa. Alla risorsa è associata una serie di numeri che rappresentano un livello astratto sopra l'URI effettivo;
  - **versionamento**: il pid è associato ad una certa risorsa in un dato momento. Se si crea una versione aggiornata della risorsa è buona pratica creare una nuova scheda con un nuovo pid;
  - **replicabilità**: attraverso il pid è possibile accedere a specifiche versioni di risorse ed eventualmente replicare un esperimento con esattamente gli stessi dati;
  - **citazione** delle risorse: il pid deve essere usato per citare correttamente una risorsa



Citare una risorsa



- Attraverso il bibtex:

```
@misc{20.500.11752/ILC-66,
 title = {{IWN}-{LOD}},
 author = {Bartolini, Roberto},
 url = {http://hdl.handle.net/20.500.11752/ILC-66},
 note = {Digital Repository for the {CLARIN} Research Infrastructure provided by {ILC}-{CNR}},
 year = {2016} }
```

- Attraverso il CMDI (VLO)

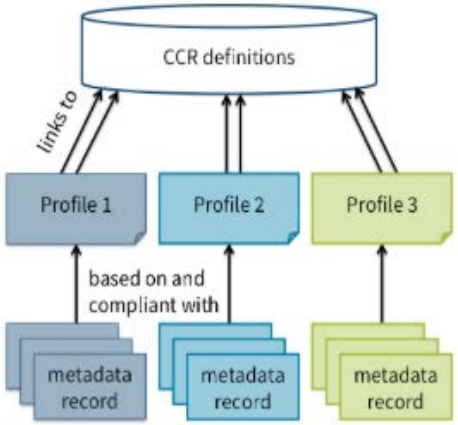
```
<cmd:CMD xmlns:cmd="http://www.clarin.eu/cmd/" xmlns:olac="http://www.clarin.eu/cmd/" xmlns:ms="http://www.clarin.eu/cmd/"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" CMDVersion="1.1" xsi:schemaLocation="http://www.clarin.eu/cmd/
http://catalog.clarin.eu/ds/ComponentRegistry/rest/registry/profiles/clarin.eu:cr1:p_1403526079380/xsd">
 <cmd:Header>
        <cmd:MdSelfLink>http://hdl.handle.net/20.500.11752/ILC-66@format=cmdi</cmd:MdSelfLink>
        .....
        <cmd:MdCollectionDisplayName>ILC Data &amp;amp; Tools</cmd:MdCollectionDisplayName>
 </cmd:Header>
 -----
</cmd:CMD>
```

- Attraverso il pid direttamente

  — **http://hdl.handle.net/20.500.11752/ILC-66**

CMDI in breve

- Acronimo di Component MetaData Infrastructure;
- Basato sulla creazione e riuso di componenti
- un componente è un set minimo di informazioni adatto a descrivere un aspetto di una risorsa:
  - componente per l'autore
  - componente per il mediatype
  - componente per il tipo della risorsa
  - componente per il contatto
  - ...
- I componenti possono essere innestati l'uno negli altri creando una gerarchia potenzialmente complessa



- Il CMDI è fondamentale per gara se descritte
- Il VLO usa il CMDI per mappare i sui propri filtri di ricerca

Il CMDI è complesso...

Il CMDI del contatto è come segue

```
<CMD_ComponentSpec isProfile="false">
 <Header>
                <ID>clarin.eu:cr1:c_1264757339280</ID>
                                <Name>component-contact
                                                <Description>I
                        </Description>
 </Header>
 <CMD_Component CardinalityMax="1" CardinalityMin="1"
   <CMD_Element CardinalityMax="1" CardinalityMin="0" ValueScheme="string"
          ConceptLink="http://www.isocat.org/datcat/DC-2454" name="Name"/>
                                                                        .....
       </CMD_Component>
     </CMD_ComponentSpec>
```



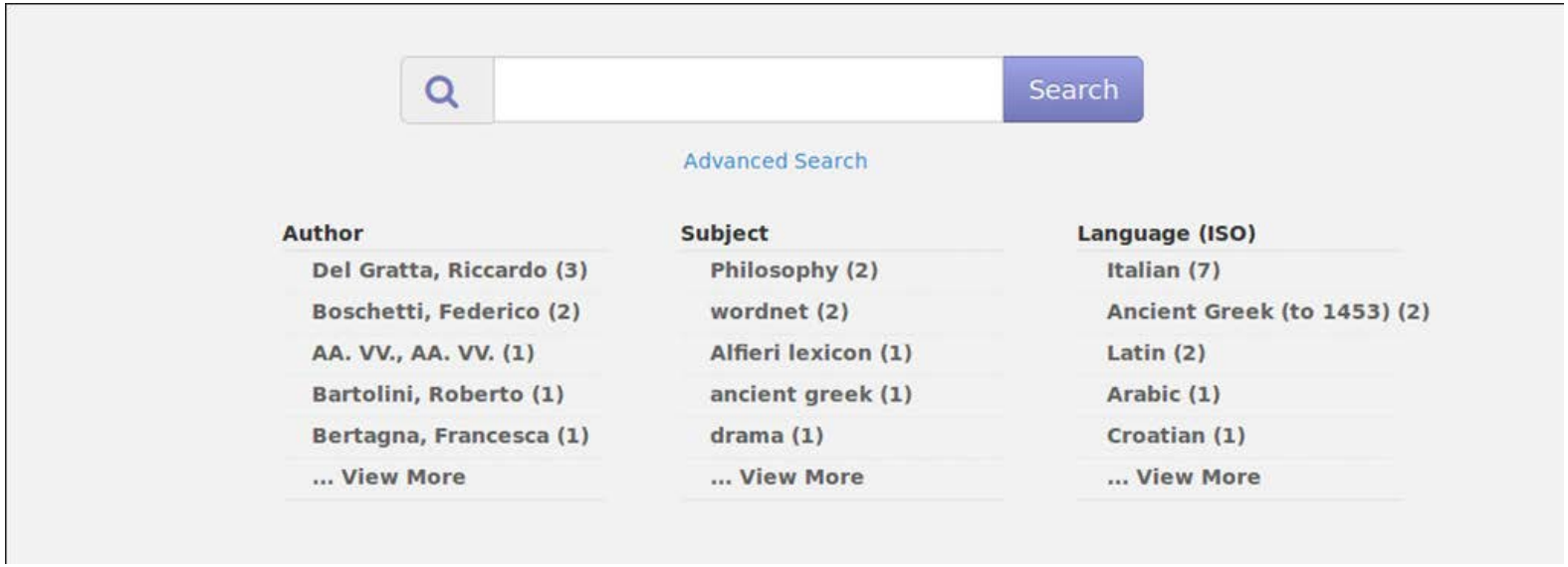
.....Ma nessun problema, clarin-dspace converte automaticamente i valori immessi dall'utente nello specifico componente

Link utili:

- https://www.clarin.eu/content/component-metadata
- https://catalog.clarin.eu/ds/ComponentRegistry

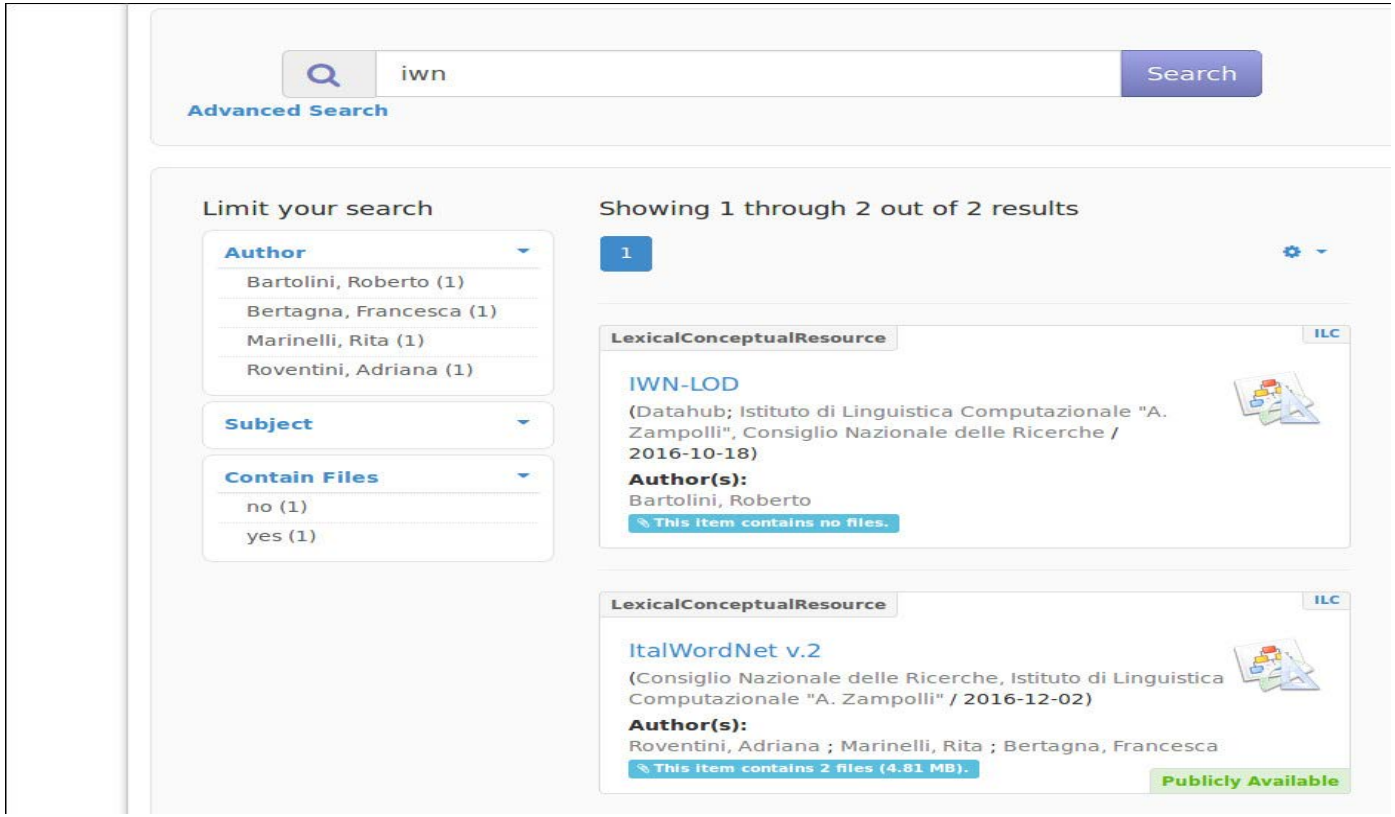
Navigare il repository

- la ricerca di schede nel repository non ha bisogno di autenticazione



Navigare il repository

- la ricerca libera può essere ristretta attraverso un sistema a faccette



Navigare il repository



Tipo di risorsa descritta

Navigare il repository



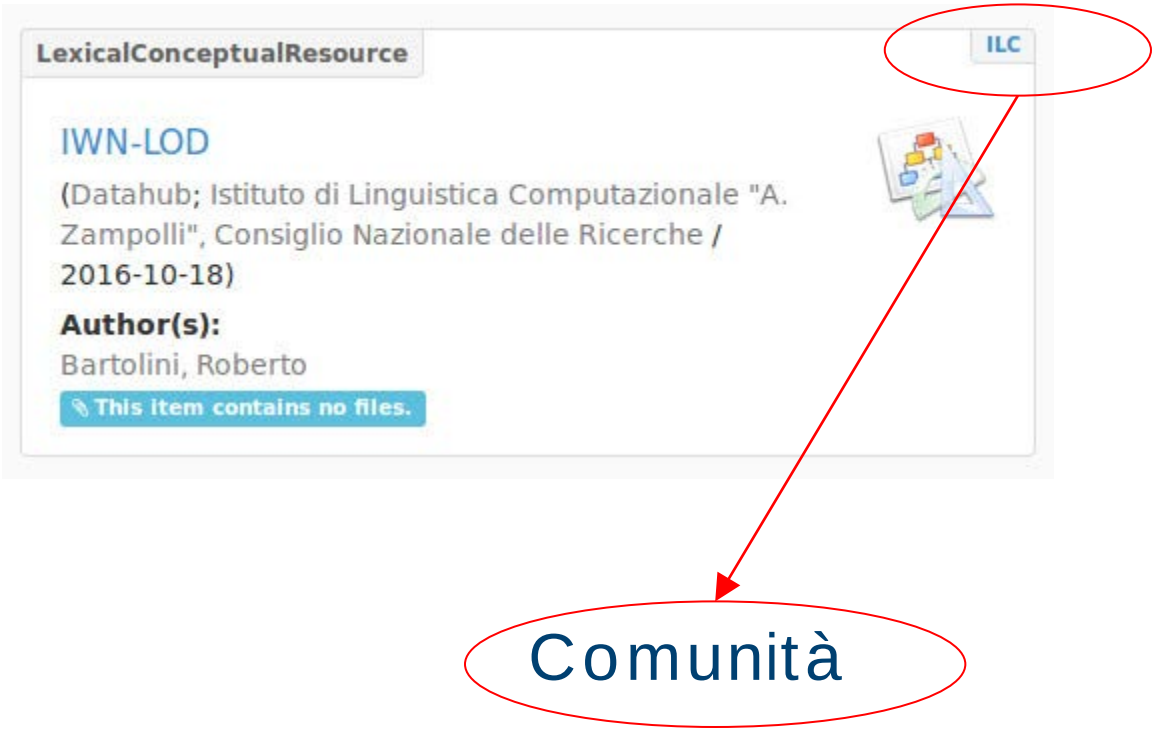
Link diretto alla scheda

Navigare il repository



Informazione sull'eventuale contenuto di file

Navigare il repository



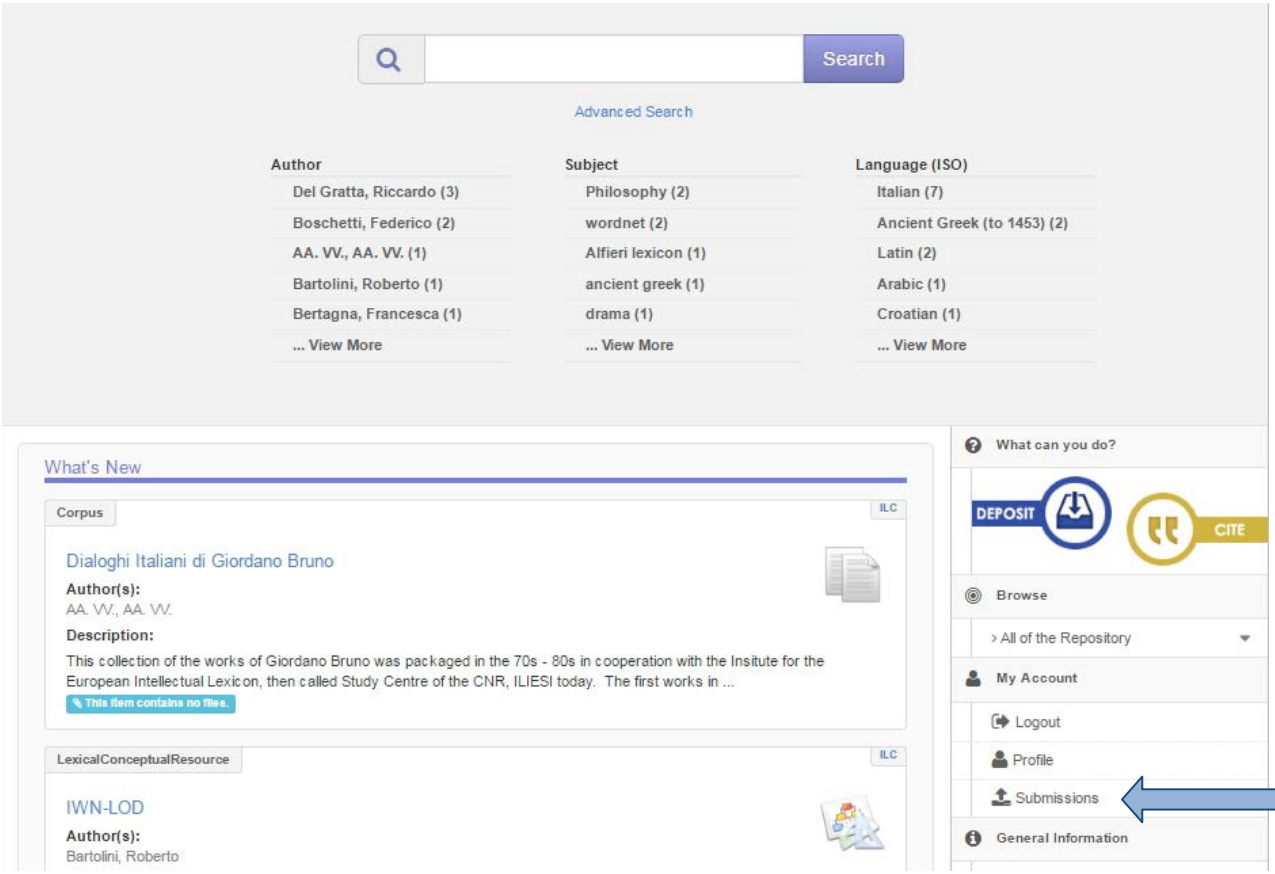
Navigare il repository



Questa scheda ha dei file connessi, e quindi necessita di opportuna licenza.

La catalogazione

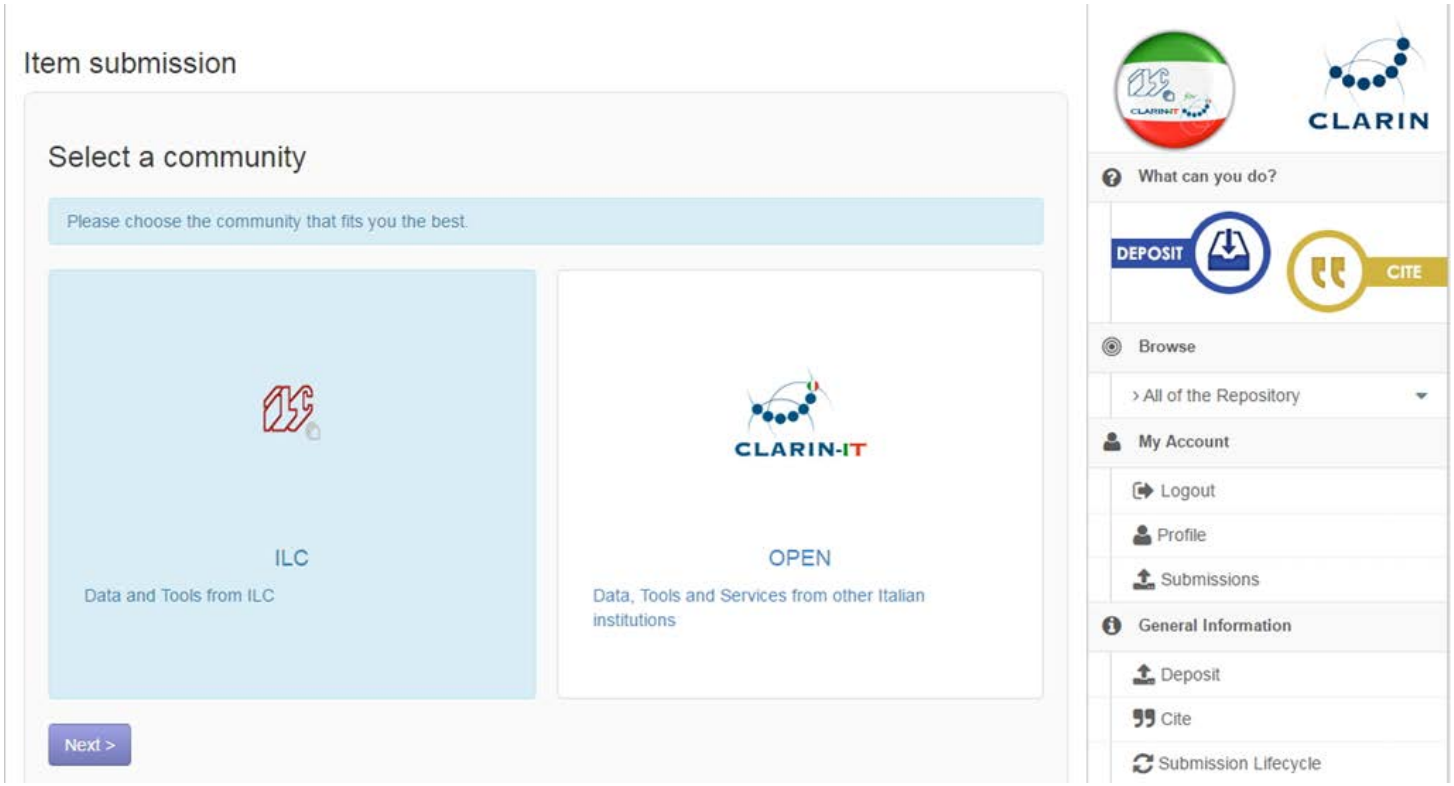
- Come inserire e descrivere una risorsa (metadati)



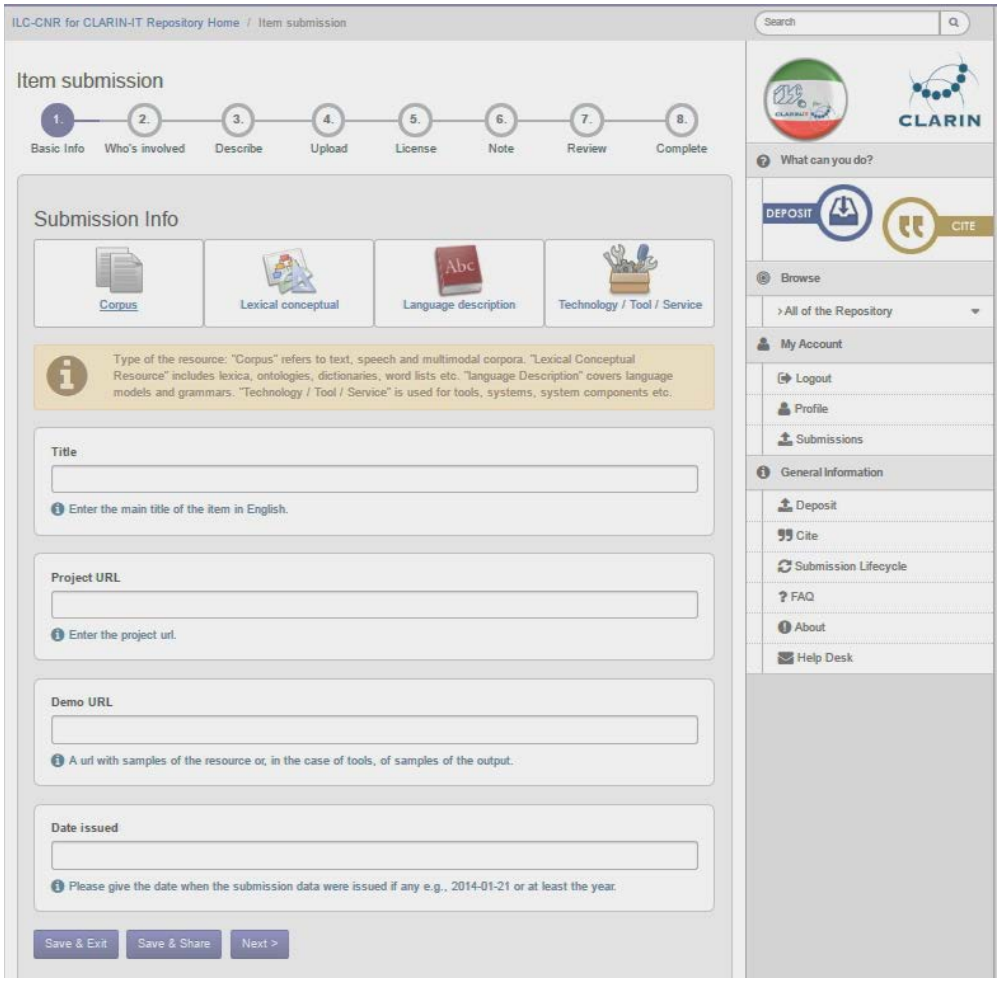
Dopo aver effettuato l'accesso con le proprie credenziali scegliere dal menù a destra "Submissions"

La catalogazione

- Si possono inserire due tipologie di dati:
- Risorse e Strumenti di ILC (ed è il caso nostro);
- Risorse, Strumenti e Servizi di altre istituzioni italiane.



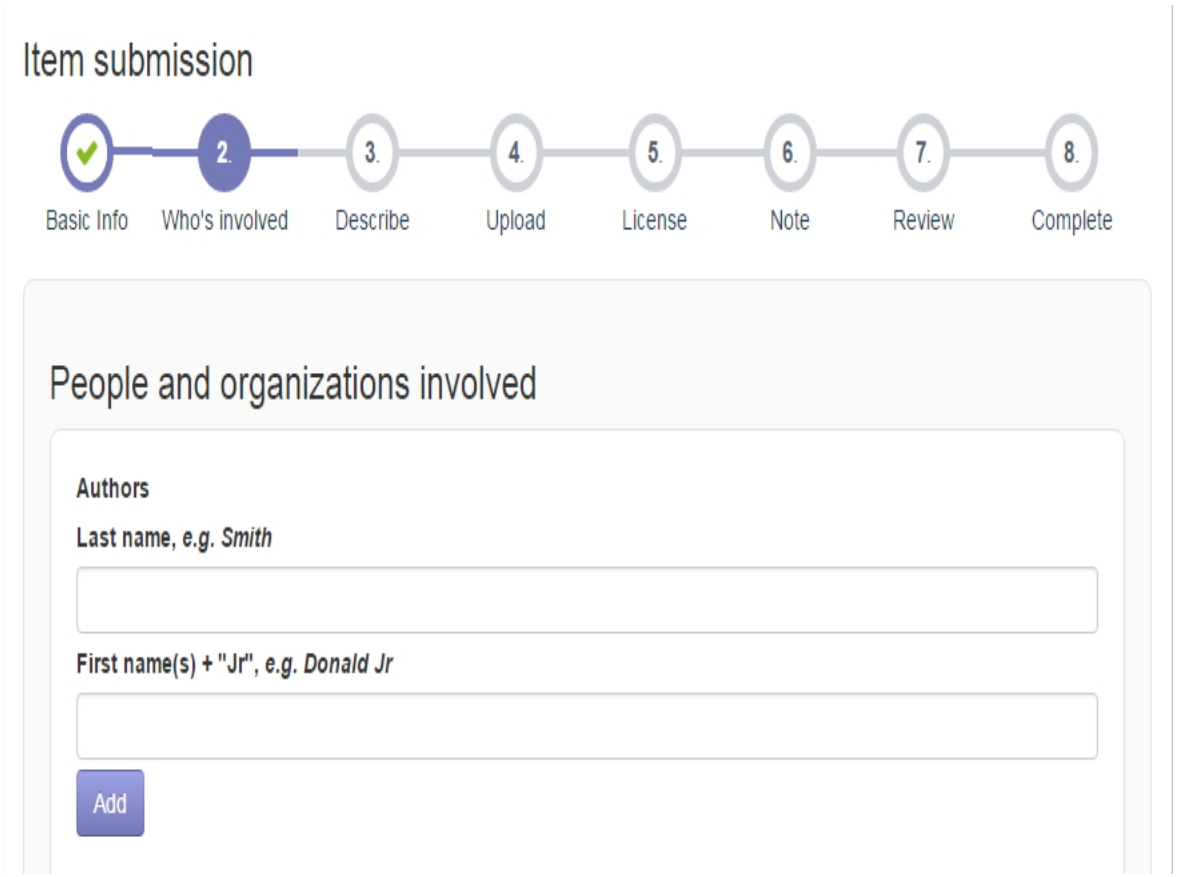
La catalogazione



- Titolo della risorsa
- URL del progetto nel caso in cui è stata realizzata nell'ambito di un progetto;
- URL della demo (se previsto)
- Data Di inserimento della risorsa in CLARIN;

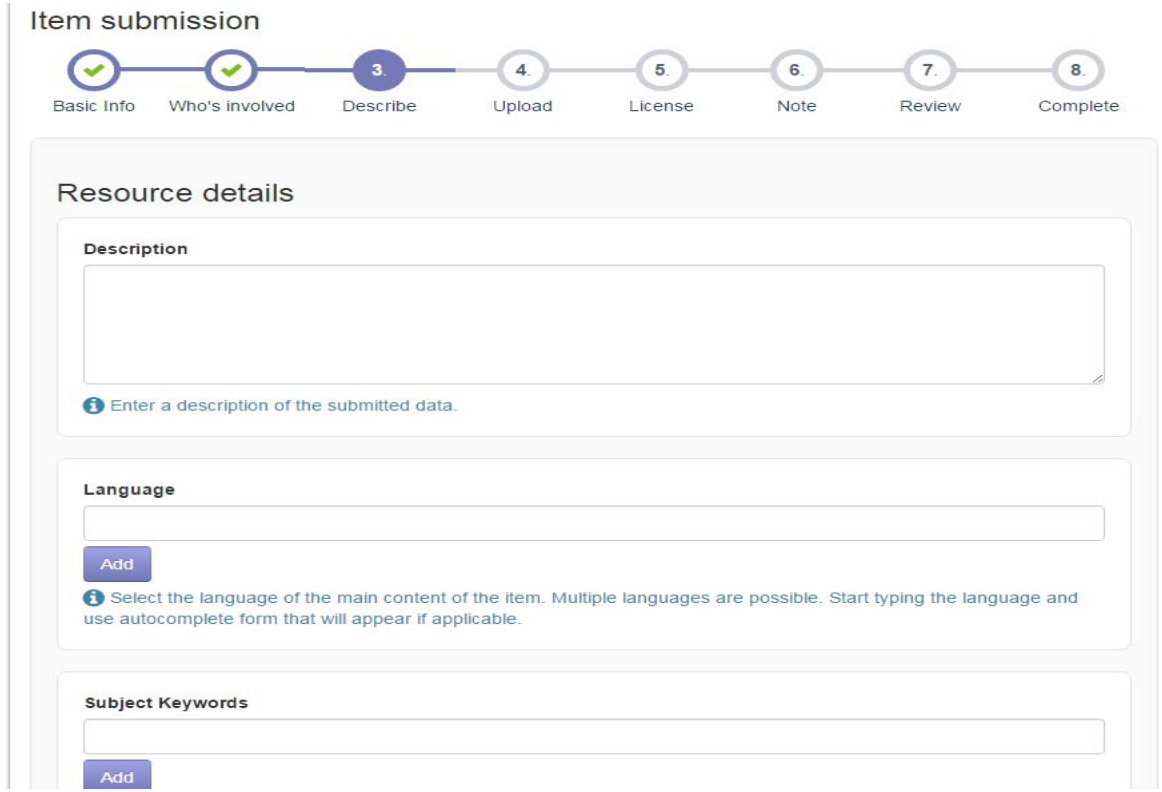
Catalogazione: descrivere una risorsa

➢ Autori



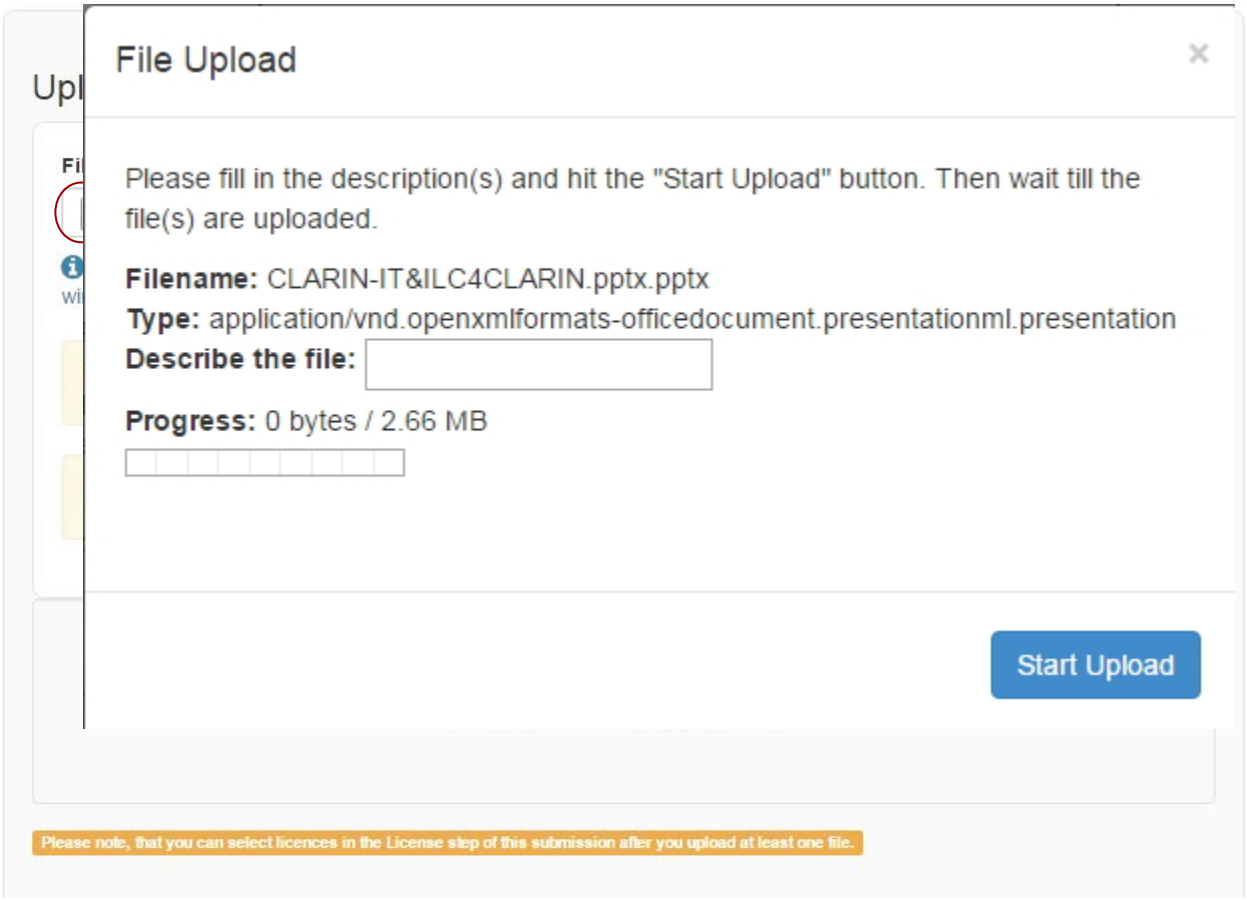
Catalogazione: descrivere una risorsa

- Descrizione - Parole chiavi -Mail delle persone che possono essere contattate



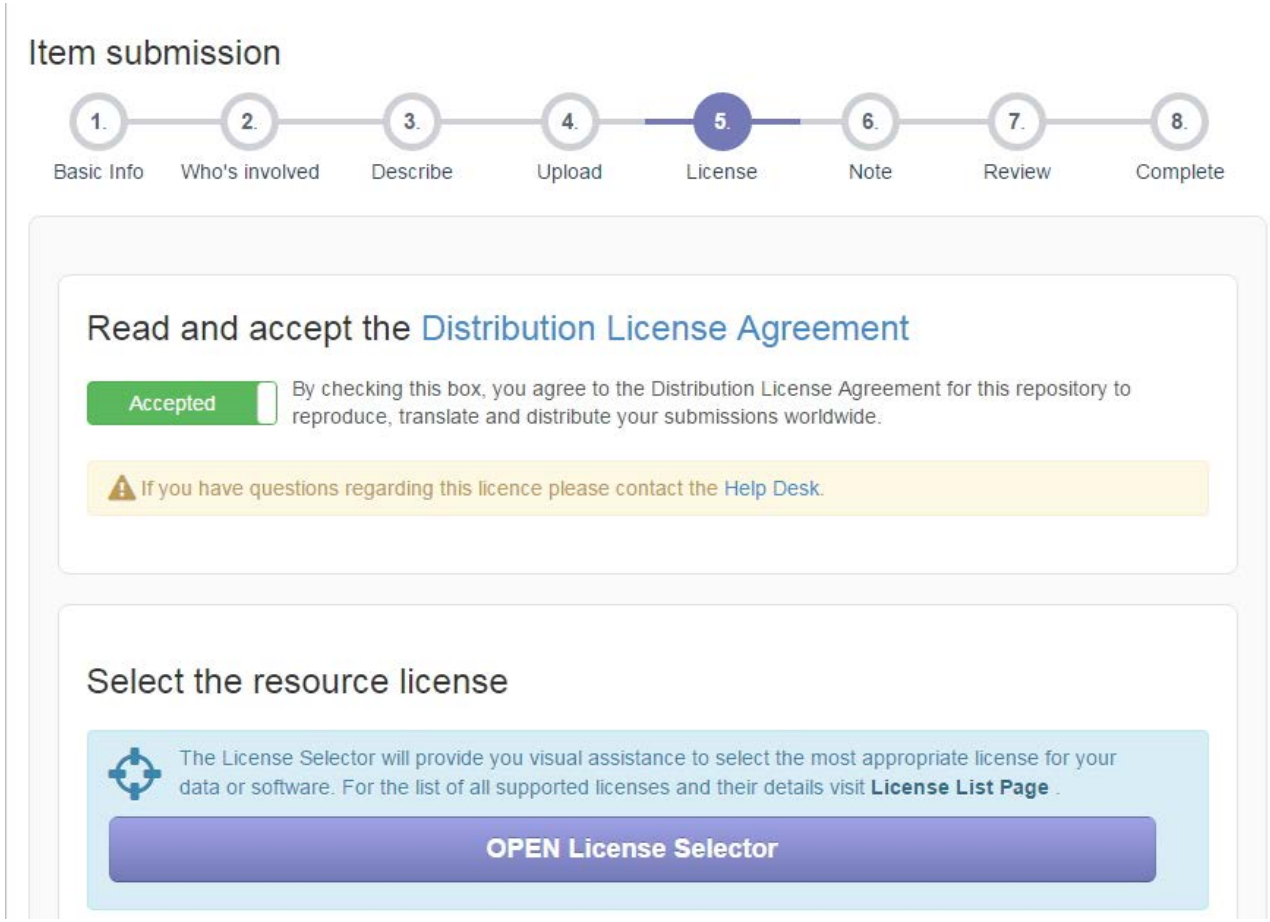
Catalogazione: descrivere una risorsa

- Upload del file

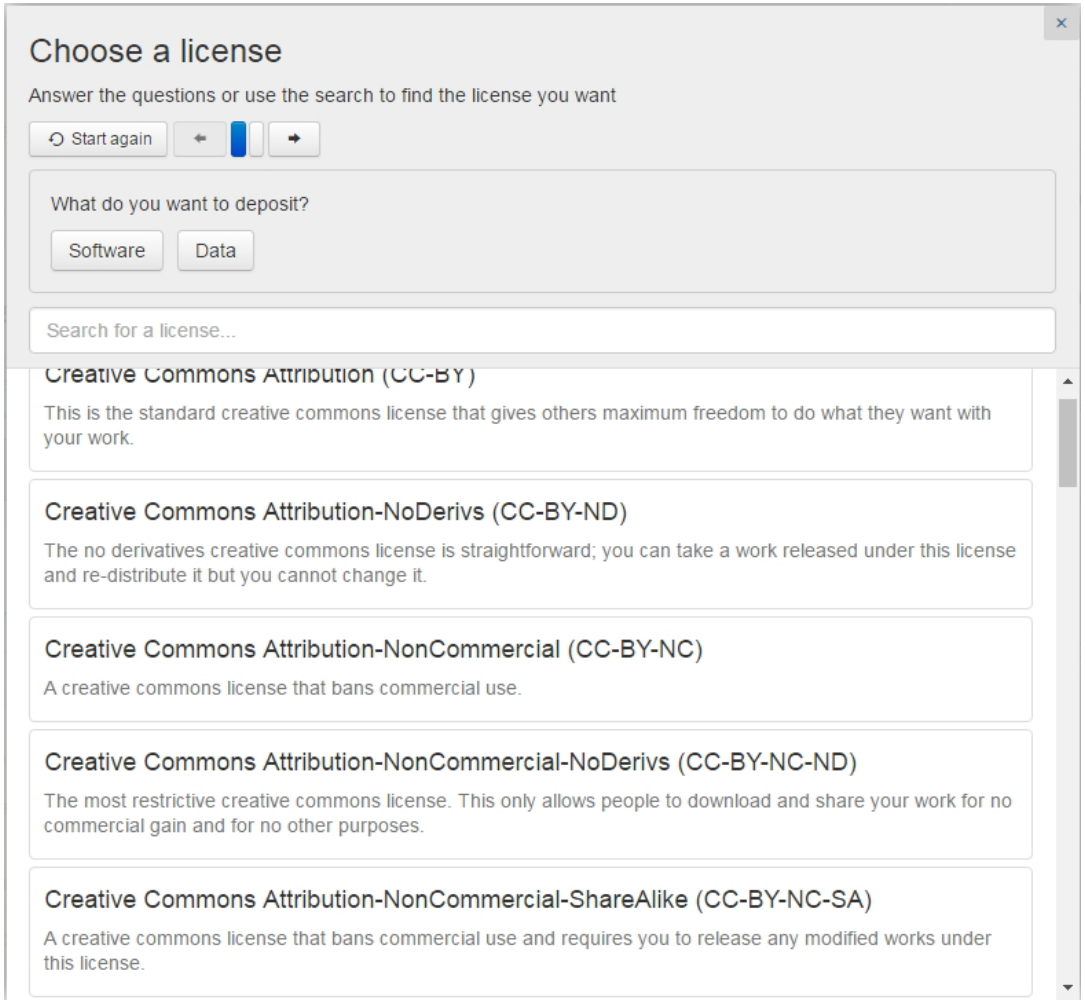


Catalogazione: descrivere una risorsa

Selezionando questa casella, l'utente accetta il contratto di licenza di distribuzione per il repository e autorizza la distribuzione e condivisione con tutti.

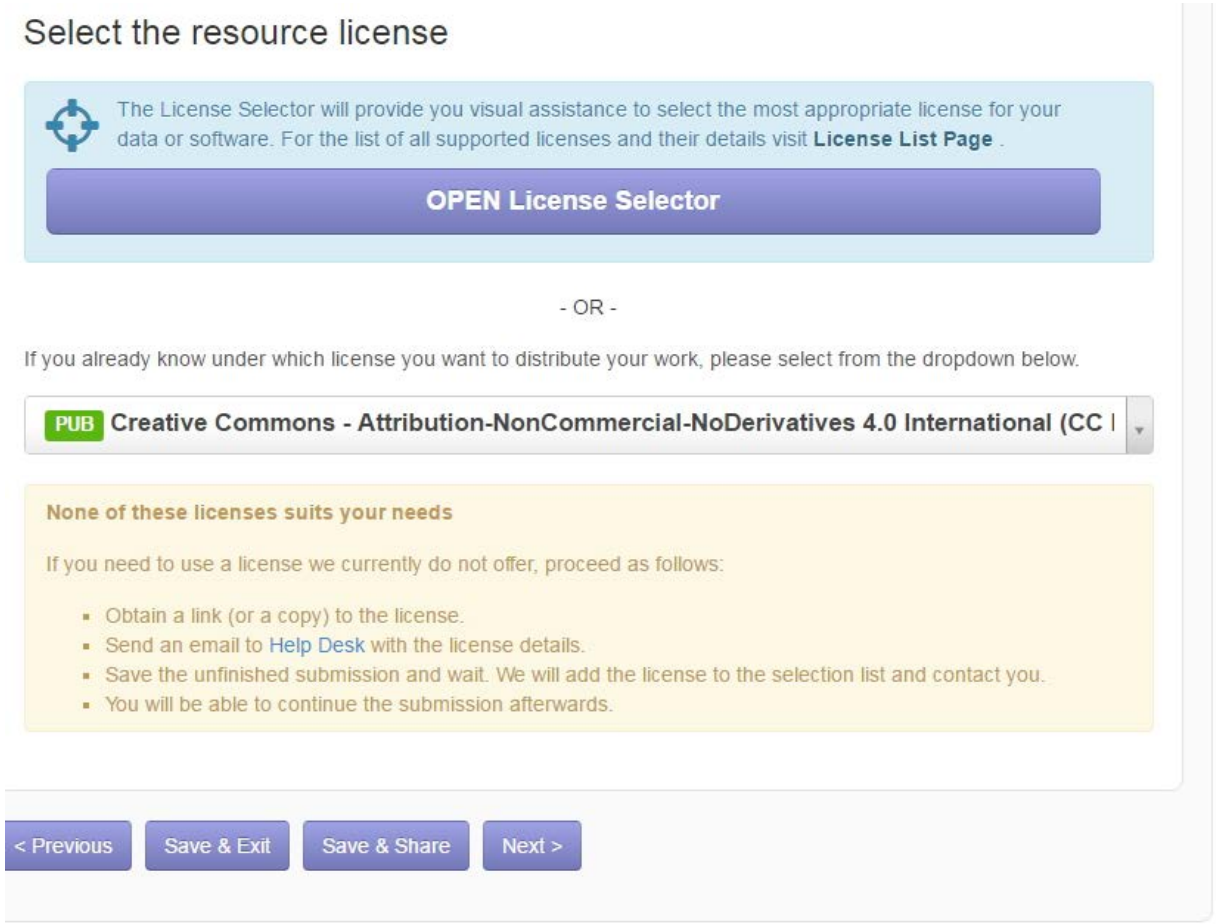


Catalogazione: descrivere una risorsa

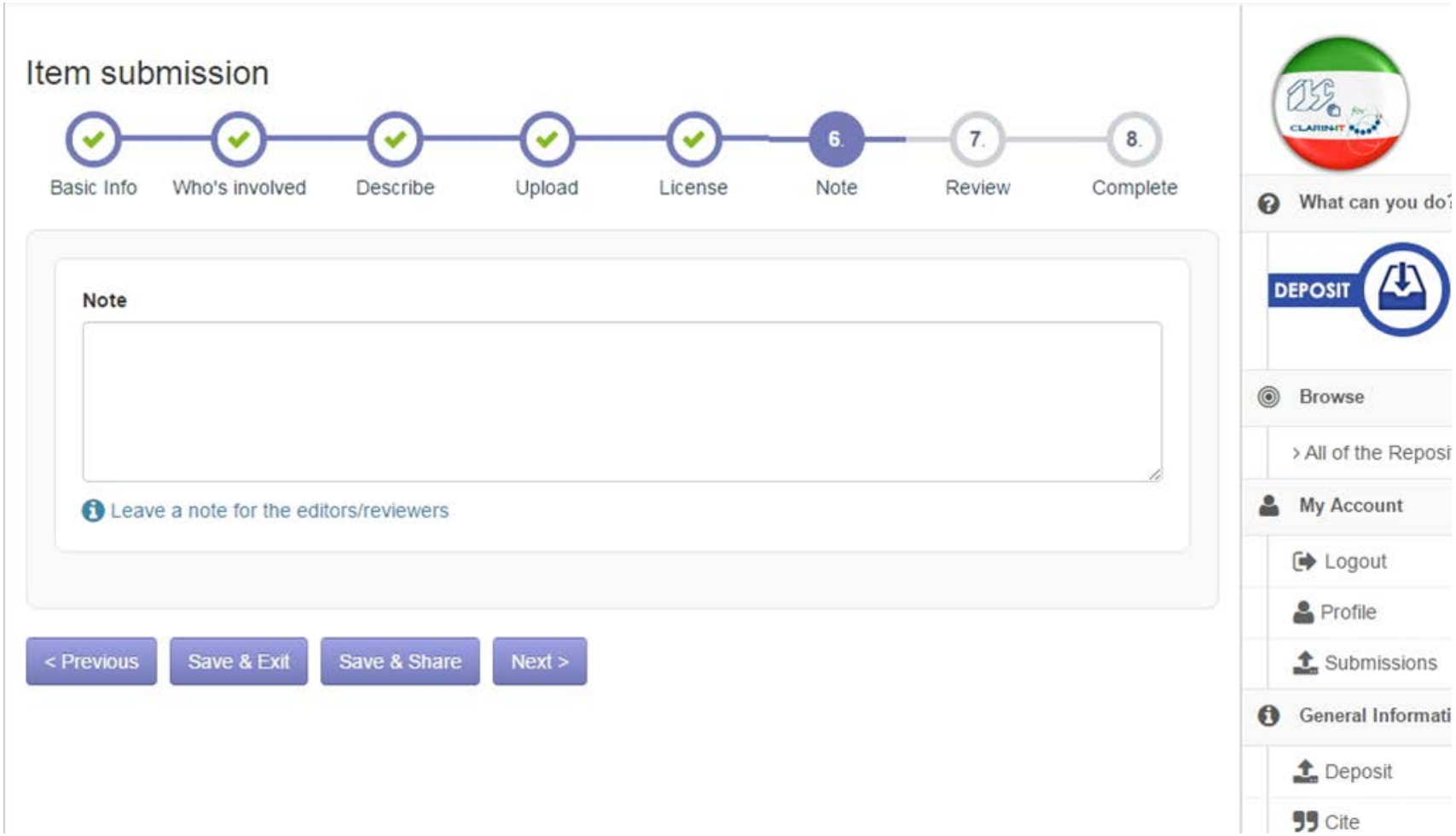


Dopo aver caricato i dati si deve scegliere una licenza. Il sito ne mette a disposizione varie tipologia in base alle esigenze dell'utente.

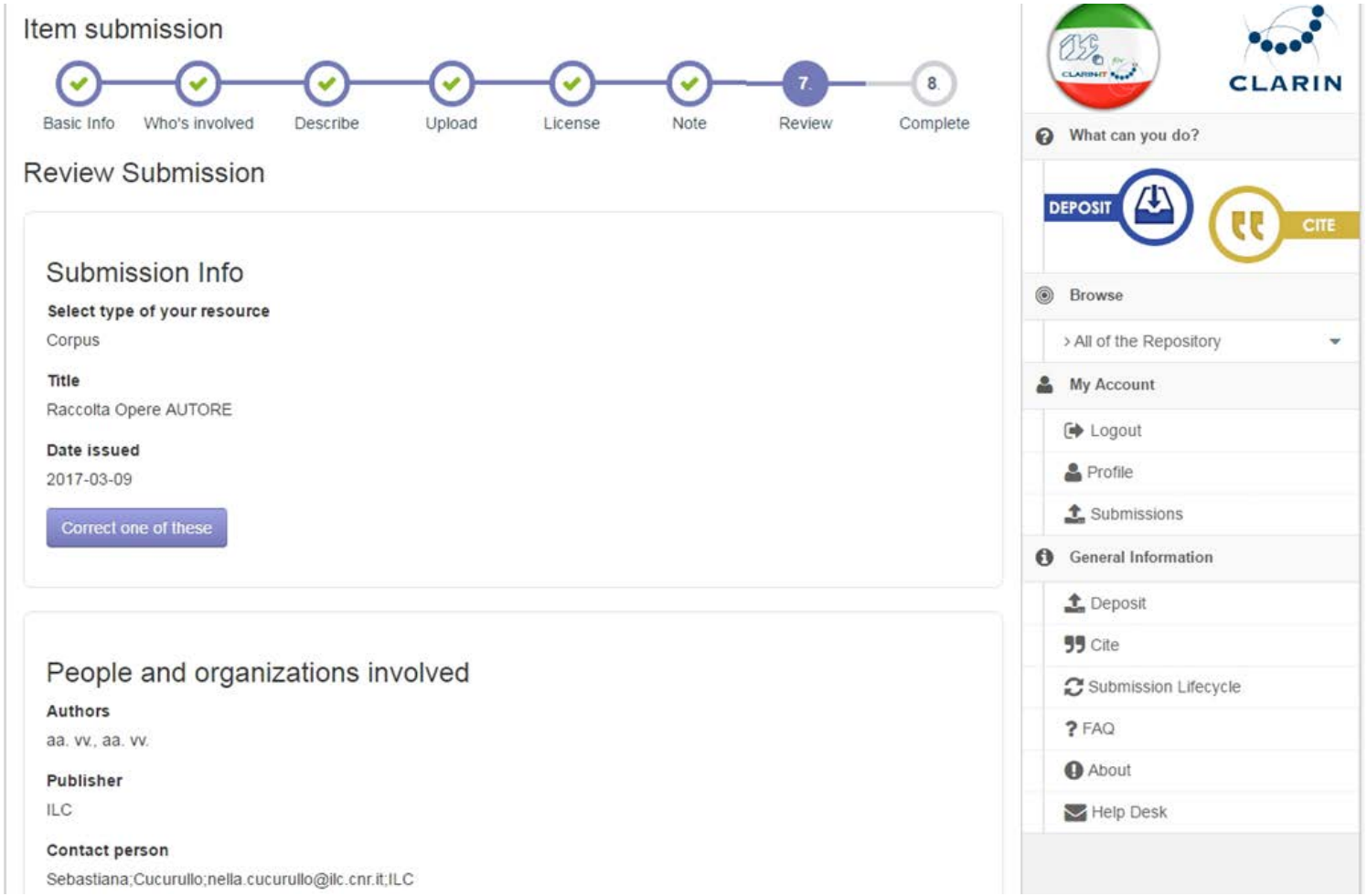
Catalogazione: descrivere una risorsa



Catalogazione: descrivere una risorsa



Catalogazione: descrivere una risorsa



Catalogazione: descrivere una risorsa



Upload File(s)

CLARIN-IT&ILC4CLARIN.pptx.pptx - Microsoft Powerpoint XML (Known)

Correct one of these

Licensing

Creative Commons - Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0)

Correct one of these

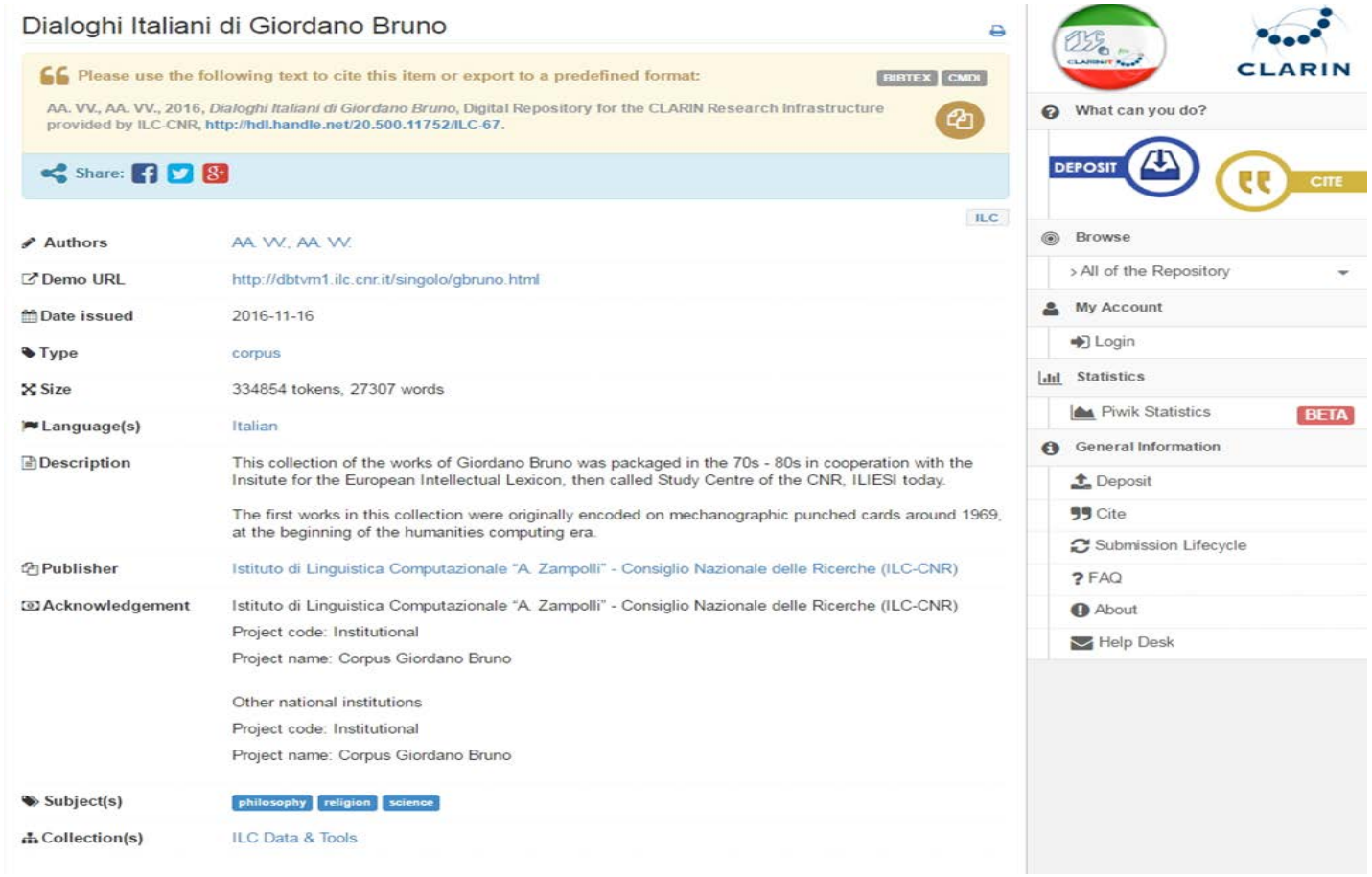
Additional Metadata for this Item

Correct one of these

< Previous | Save & Exit | Save & Share | Complete submission

La catalogazione

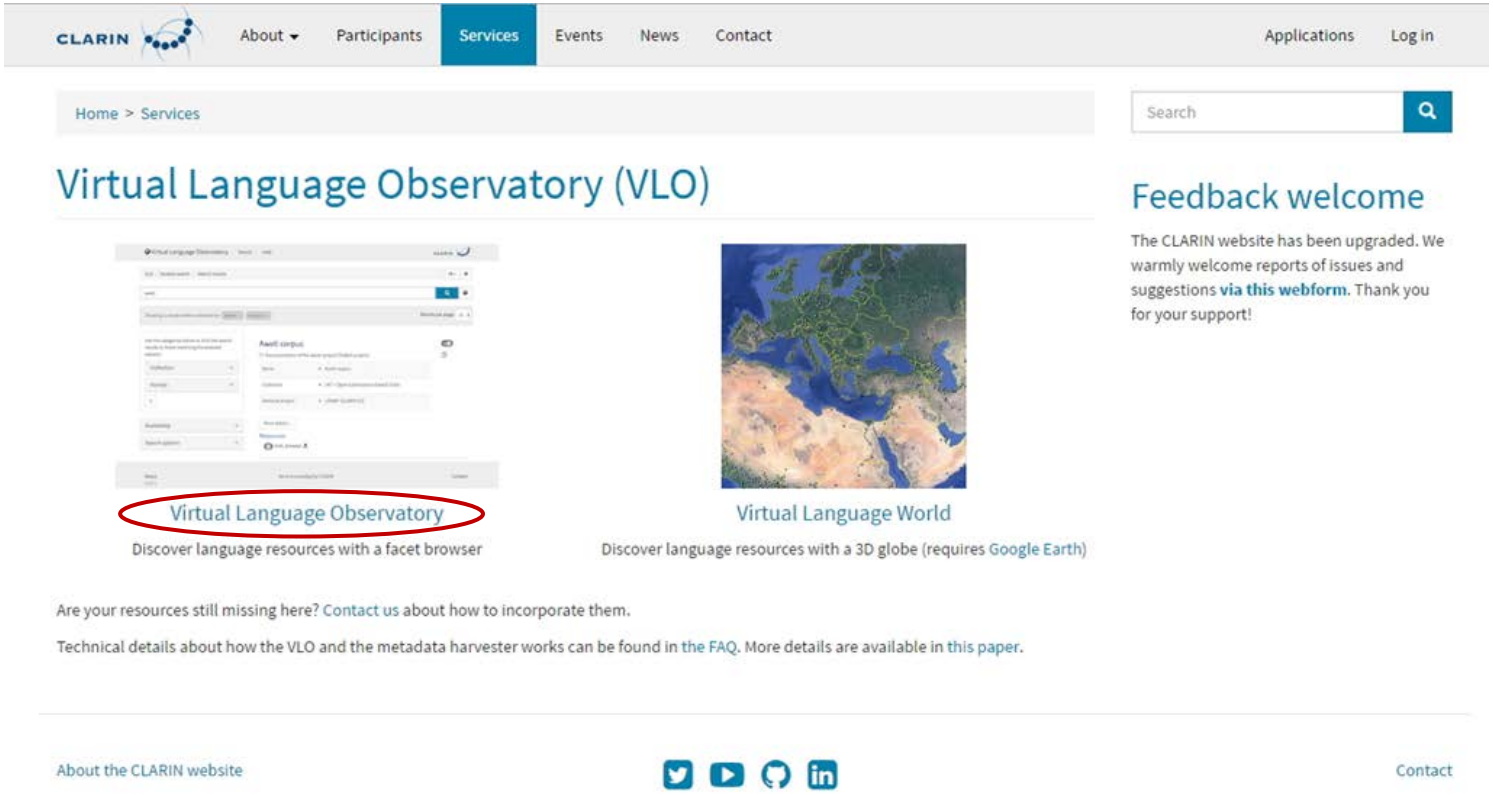
- Esempio



Virtual Language Observatory
VLO

https://www.clarin.eu/content/virtual-language-observatory-vlo

- VLO (Virtual Language Observatory): aiuta i ricercatori a trovare facilmente le risorse linguistiche e gli strumenti per il loro lavoro di ricerca.



Virtual Language Observatory
VLO



Virtual Language Observatory
VLO



Virtual Language Observatory Search Help CLARIN

VLO / Faceted search / Search results

Giordano Bruno

Showing 2 results for Giordano Bruno ×

Results per page: 10

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

- Language
- Collection
- Resource type
- Format

Availability

Search options

Dialoghi Italiani di Giordano Bruno

This collection of the works of Giordano Bruno was packaged in the 70s - 80s in cooperation with the Insitute for the European Intellectual Lexicon, then called Study Centre of the CNR, ILIESI today. The first works in this collection were originally encoded on mechanographic punched cards around 1969, at the beginnin...

DE VRIJE GEDACHTE

Serie programma's over 'De vrije gedachte', de stem van vrijdenkend Nederland, met in deze aflevering vanaf verschillende locaties in Rome een overdenking van Paul Cliteur, voorzitter Humanistisch Verbond, over wetenschap en religie. Aan de orde komen: de betekenis van religie; verschillen en overeenkomsten van religie...

Virtual Language Observatory
VLO



Virtual Language Observatory Search Help CLARIN

VLO / Faceted search / Search results / Record: Dialoghi Italiani di Giordano Bruno

Record 1 of 2 < previous next >

Dialoghi Italiani di Giordano Bruno

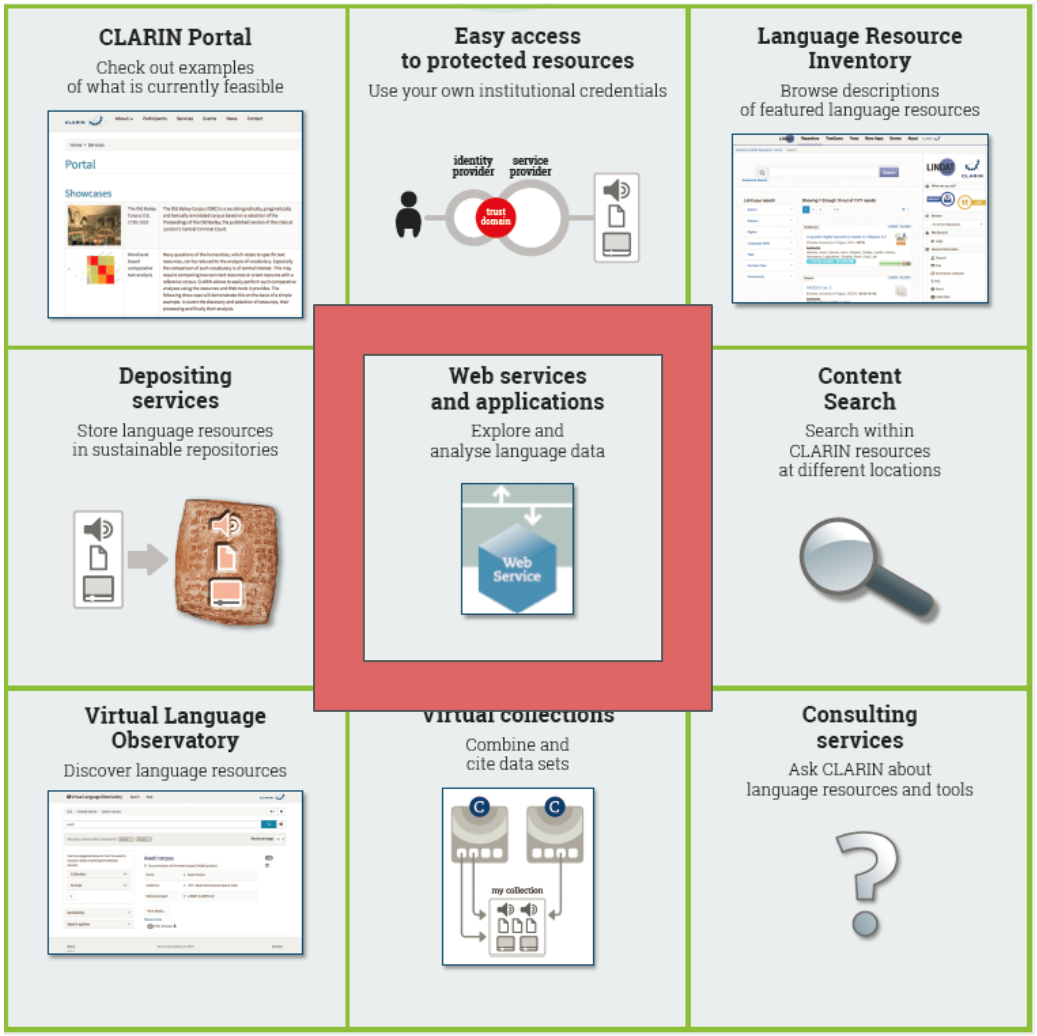
Show the original provider's page for this record

Record details | Resources (0) | Availability | All metadata | Technical details

Name	Dialoghi Italiani di Giordano Bruno
Description	This collection of the works of Giordano Bruno was packaged in the 70s - 80s in cooperation with the Insitute for the European Intellectual Lexicon, then called Study Centre of the CNR, ILIESI today. The first works in this collection were originally encoded on mechanographic punched cards around 1969, at the beginning of the humanities computing era.
Collection	ILC Data & Tools
Language	Italian
Organisation	Istituto di Linguistica Computazionale "A. Zampolli" - Consiglio Nazionale delle Ricerche (ILC-CNR)
National project	CLARIN-IT

About v4.0.2 Service provided by CLARIN Contact

I servizi: sommario



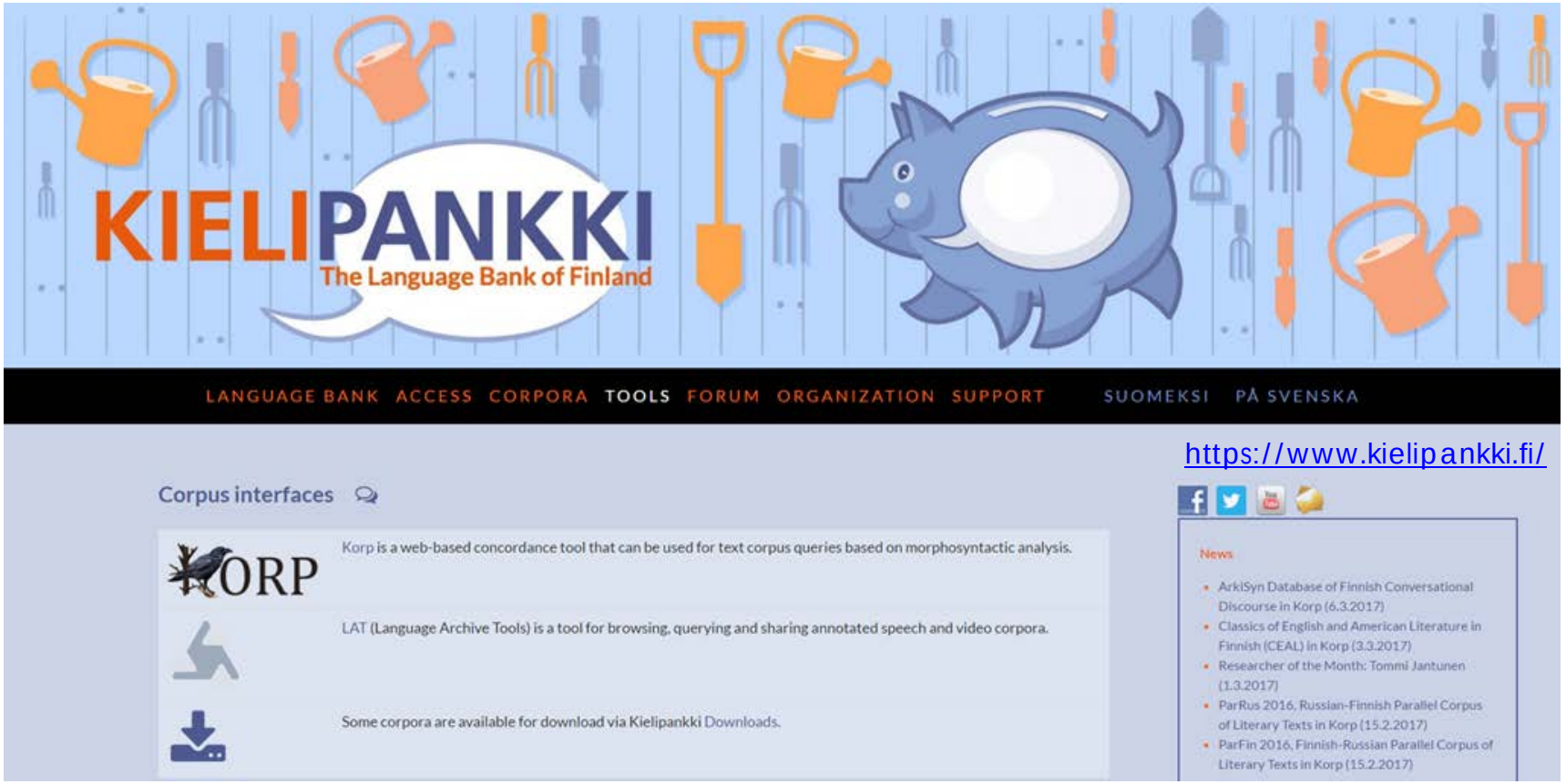
Servizi linguistici dei centri nazionali



- Accesso e/o consultazione a corpora, treebanks, lessici, ...
- Servizi di analisi e annotazione linguistica
- Accesso alla Ricerca di contenuto federata
- Applicazioni più complesse per le SSH

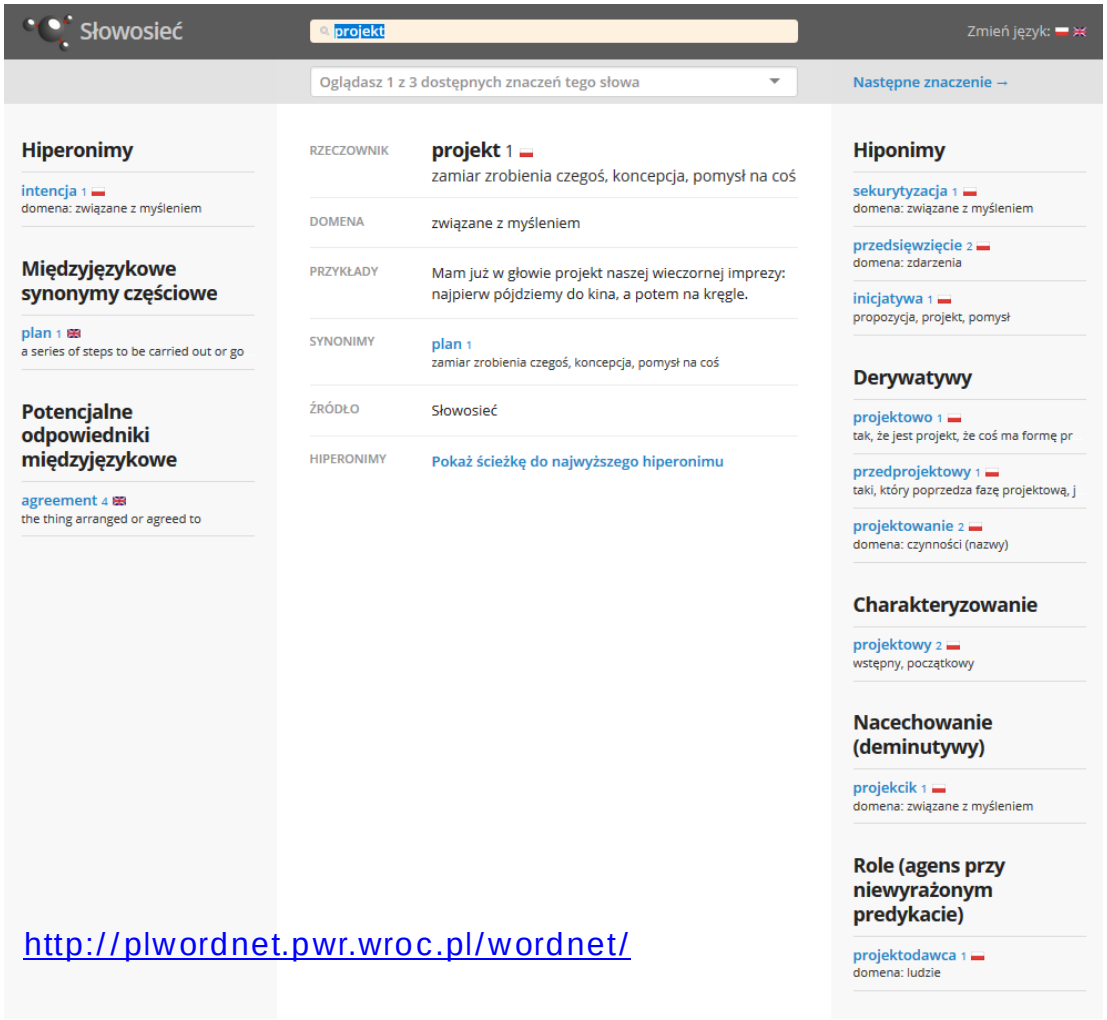
--> CLARIN-ERIC consente di riusare in CLARIN strumenti già esistenti

Esempi dagli altri centri: concordanze e multimodalità



pochi, semplici servizi simpaticamente confezionati

Esempi dagli altri centri: interrogazione WordNets



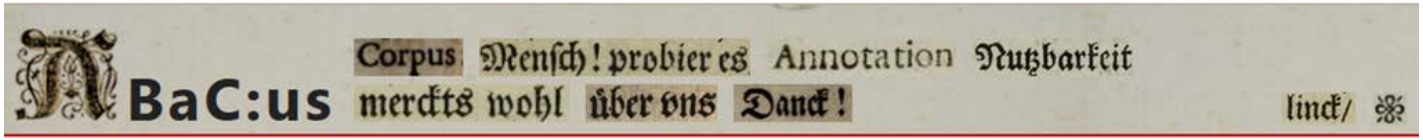
http://plwordnet.pwr.wroc.pl/wordnet/

Esempi dagli altri centri: catena di analisi linguistica



https://clarin-pl.eu/dspace/handle/11321/7

Esempi dagli altri centri: consultazione testi antichi



Geneigter Leser.

Herzlich willkommen auf den Seiten des Austrian Baroque Corpus!



Aber was ist das?

ABaC:us – Austrian Baroque Corpus ist die Bezeichnung für eine digitale Sammlung von literarischen Texten aus der Barockzeit. Die Texte, die Sie hier einsehen und durchsuchen können, sind Teil einer reichhaltigen und weit verbreiteten Literatur, die sich mit der Vorbereitung auf den Tod beschäftigt und heute beinahe in Vergessenheit geraten ist. Den thematischen Kern der Sammlung bilden Schriften theologisch-erbaulichen Inhalts, insbesondere Ars moriendi-, Totentanz- und Memento mori-Literatur, die einen wesentlichen Teil damaliger Textproduktion darstellt.

was man Gutes gethan

Mit ABaC:us wird der Versuch unternommen, zeitentfernte Printprodukte der Barockzeit für ein interessiertes, öffentliches (Fach-) Publikum mit innovativen texttechnologischen Methoden zu erschließen. Durch die Anwendung international anerkannter Standards stellt das Projektteam sicher, dass die aufbereiteten, annotierten historischen Textdaten verschiedenen Erkenntnisinteressen dienlich sind.

In der texttechnologischen Erschließung dieser wertvollen Quellen wird sprachlich-kulturelles Erbe aus der Barockzeit bewahrt und eine empirische Grundlage zu dessen nachhaltiger, interdisziplinärer Erforschung und Interpretation geschaffen.

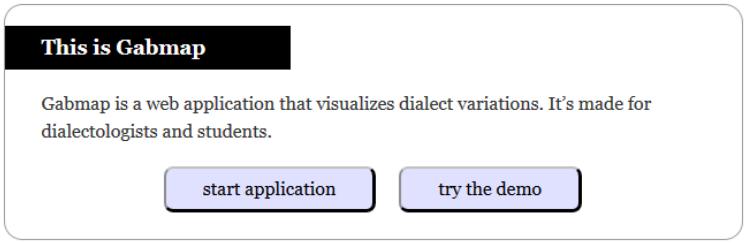
zur Welt hinaus.

Das **Austrian Centre for Digital Humanities**, das mit 1. Jänner 2015 von der **Österreichischen Akademie der Wissenschaften** in Wien gegründet worden ist, stellt einen Teil der Sammlung nun erstmals testweise, kostenfrei und in hochwertiger Qualität für den wissenschaftlichen Bedarf zur Verfügung.

https://acdh.oeaw.ac.at/abacus/

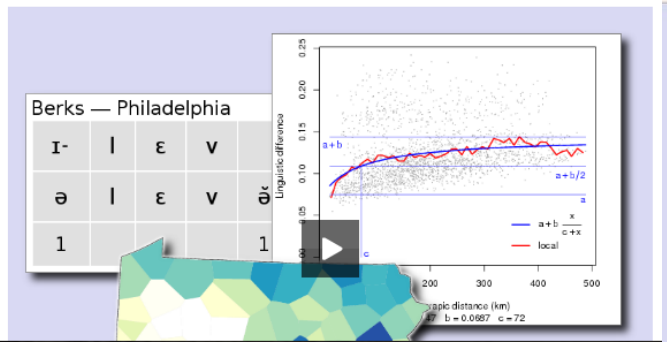
Esempi dagli altri centri: analisi dialettometrica

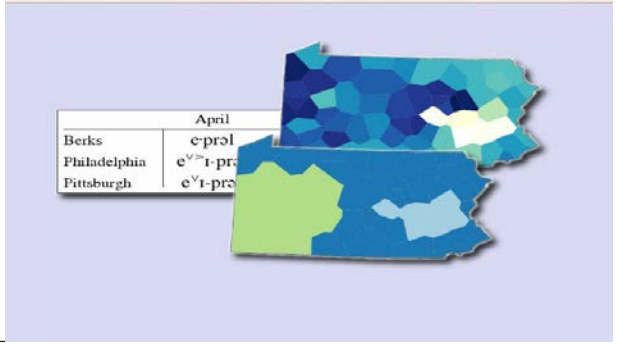




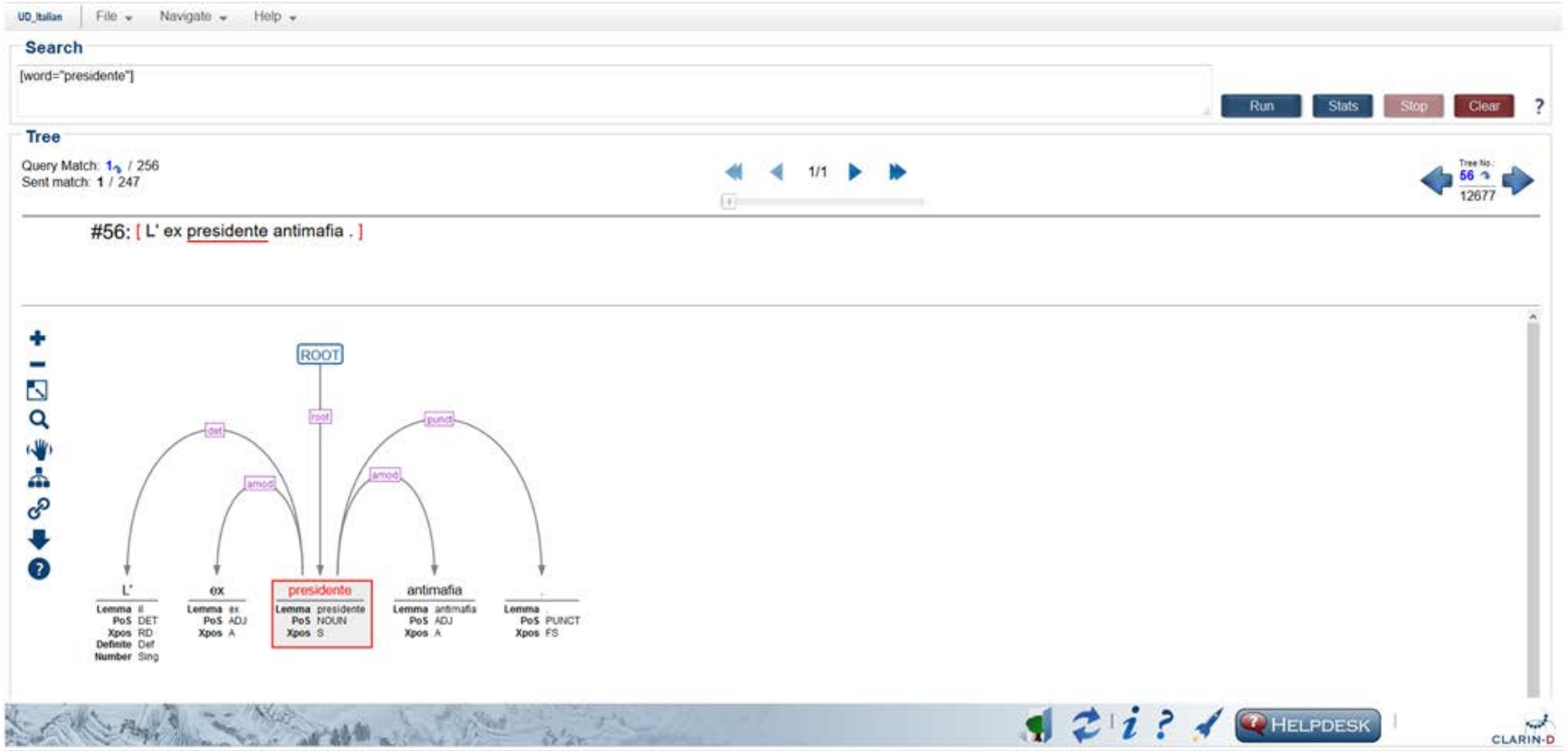
http://www.gabmap.nl

http://www.gabmap.nl/~app/doc/IntroVideo/



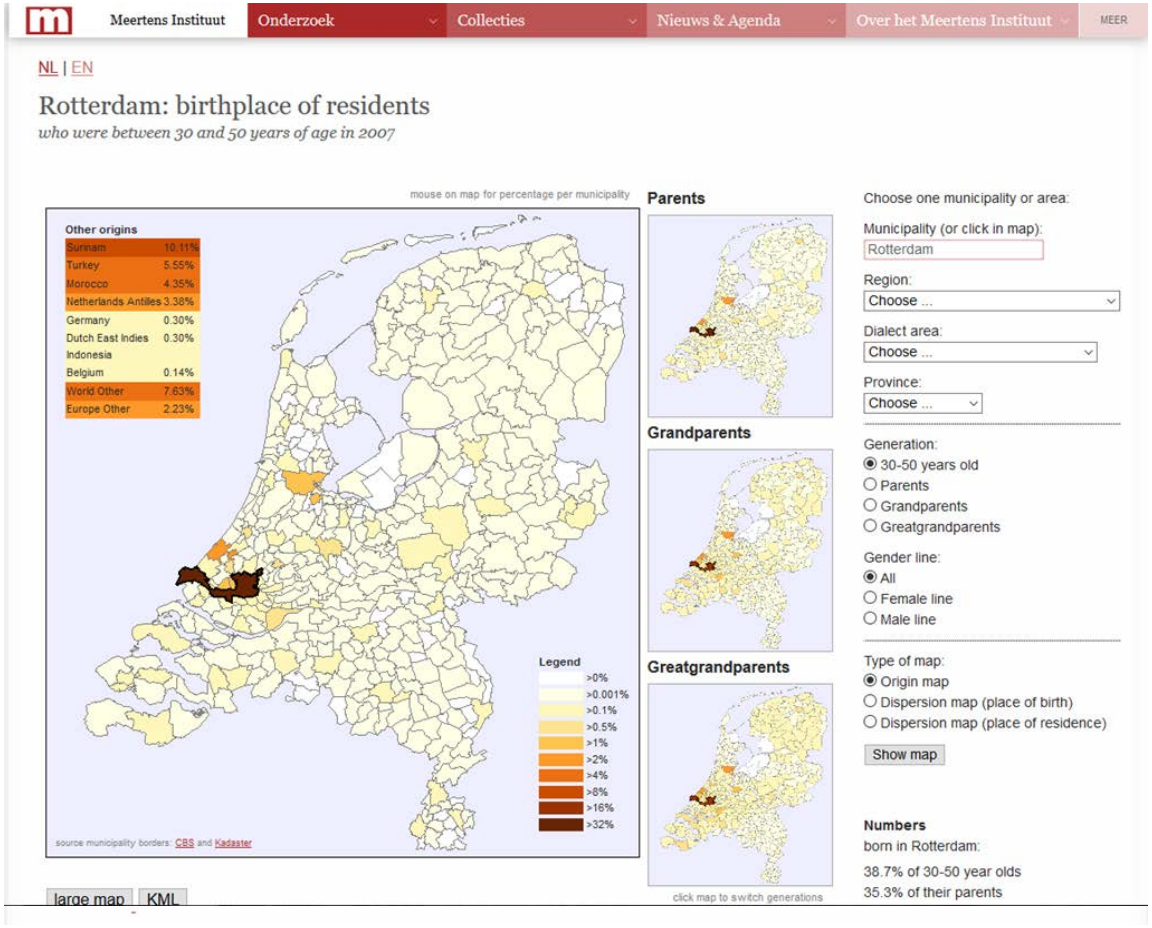


Esempi dagli altri centri: interrogazione treebanks



http://weblicht.sfs.uni-tuebingen.de/Tundra/

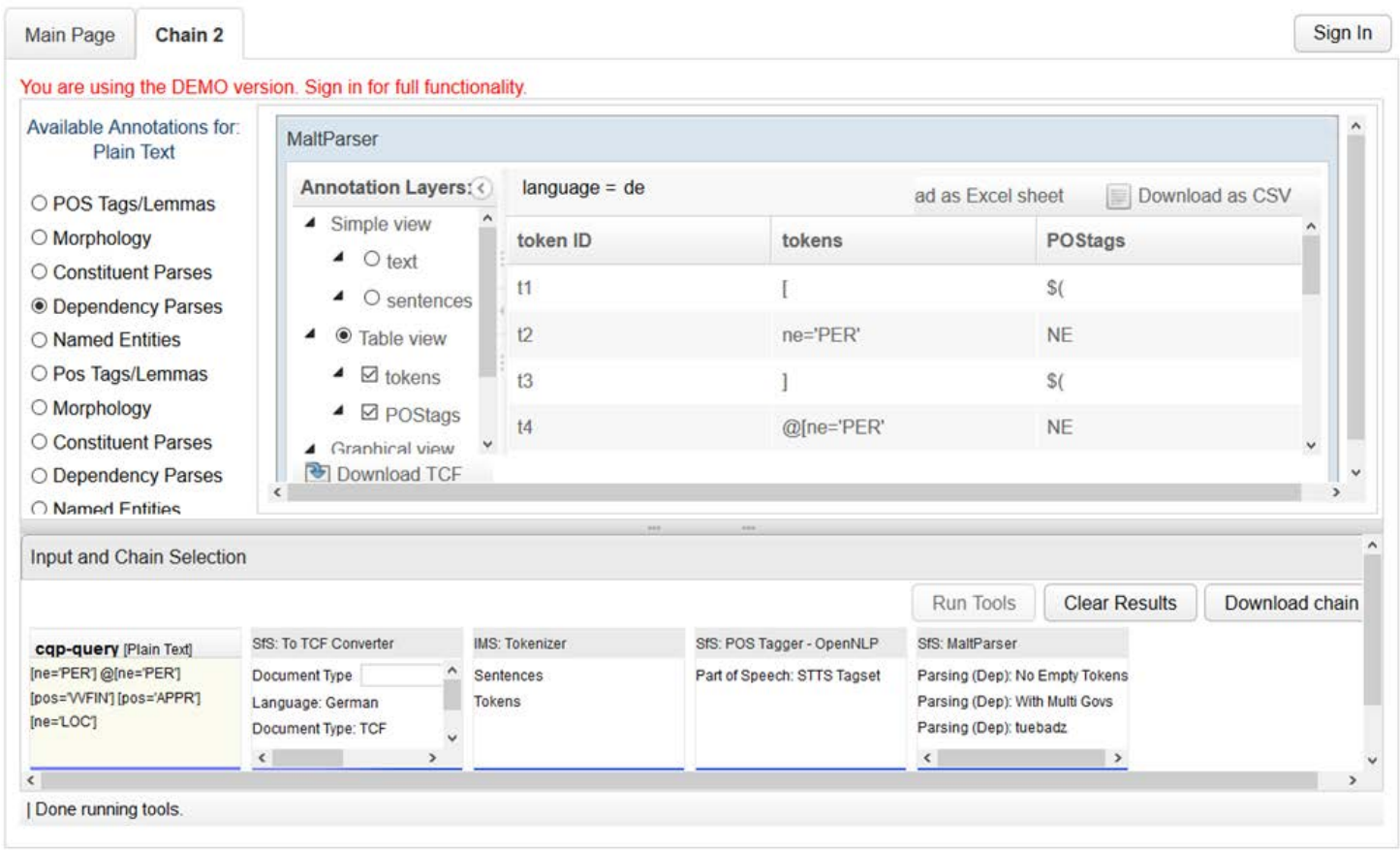
Esempi dagli altri centri: migrazioni



http://www.meertens.knaw.nl/migmap/?lang=en#

Weblicht

http://www.clarin-d.de/en/language-resources-and-services/weblicht
http://weblicht.sfs.uni-tuebingen.de/weblichtwiki/index.php/Getting_Started



https://weblicht.sfs.uni-tuebingen.de/weblicht/

Altri esempi



- Vetrina CLARIN, showcases: https://www.clarin.eu/portal

- Huygens ING (NL) https://www.huygens.knaw.nl/?lang=en
- Language Resource Portal (Vienna)https://clarin.oeaw.ac.at/ccv
- LINDAT services https://lindat.mff.cuni.cz/en/services
- CLARIN-PL http://clarin-pl.eu/en/services
- CLARIN.SI http://www.clarin.si/info/tools-and-services

Servizi ILC:
obiettivi a breve termine



Massima resa , minimo sforzo
--> Mettere a disposizione e catalogare tutto ciò che è ... esistente e operativo, 'as is' o con minimo adattamento.

- Servizi di consultazione/interrogazione/browsing di risorse (testuali, lessicali, ontologiche, multimodali ...)
- Servizi di analisi linguistica: tokenizzatori, tagger, parser, estrattori di conoscenza, visualizzatori, ...
- Applicazioni complesse, catene di analisi, ...
- Altre applicazioni web ritenute utili e rilevanti per SSH

Servizi ILC: interrogazione corpora testuali



Dialoghi italiani di Giordano Bruno

La raccolta delle opere di Giordano Bruno, è stata elaborata negli anni '70 - '80 in collaborazione con il Lessico Intellettuale Europeo, allora Centro di Studio del CNR oggi ILIESI. Le prime opere oggetto della collaborazione sono state "perf

Interrogazione Testi

Il candelaio
A cura di V. Spampanato, Laterza, 1
Ristampati con note da G. Gentile
Terza Edizione a cura di G. Aquilecchia, Sanson

Cabala del cavallo pegaseo
Dialoghi morali.
Ristampati con note da Giovanni Gentile.
Terza edizione a cura di Giovanni Aquilecchia
Sansoni, Firenze, 1958 [seconda ristampa: 1985], 833-923.

Spaccio de la bestia trionfante
Dialoghi morali.
Ristampati con note da Giovanni Gentile.
Terza edizione a cura di Giovanni Aquilecchia
Sansoni, Firenze, 1958 [seconda ristampa: 1985].

De gli eroici furori
Dialoghi morali.
Nuovamente ristampati con note da Giovanni Gentile.

I.L.C. PiSystem - Mozilla Firefox

dbtvm1.ilc.cnr.it/singolo/DBT_Isapi.dll?AZIONE=SceltaTesto&PARAM=2&LINGUA=ITA&TESTO=BR2

PiSYSTEM DBT Data Base Testuale

Param
HardFam
Guida

G. Bruno Dialoghi italiani. Dialoghi morali e dialoghi metafisici

Occorrenze :	334854	Numero totale di parole nel testo
N.Forme :	27307	Numero di forme diverse nel testo

Parole	Ricerca parole nel testo
Morfologia	Ricerca nel testo con la morfologia
Thesaurus	Ricerca nel testo con thesaurus
Punteggiatura	Ricerca della punteggiatura nel testo
Pagine	Ricerca delle pagine nel testo

DBT Picchi Eugenio. G. Bruno Dialoghi italiani. Dialoghi morali e dialoghi metafisici

Servizi ILC:
accesso a risorse lessicali

ItalWordNet

Parola: Invia Mostra tutte le relazioni

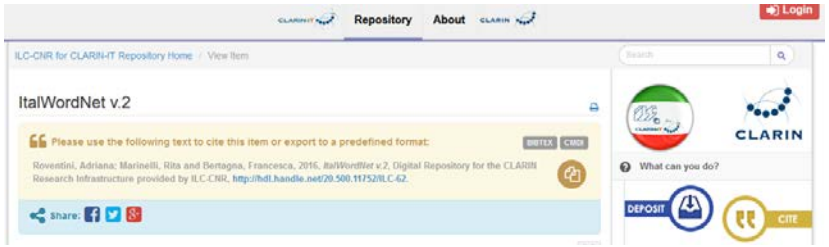
---

casa, Nome

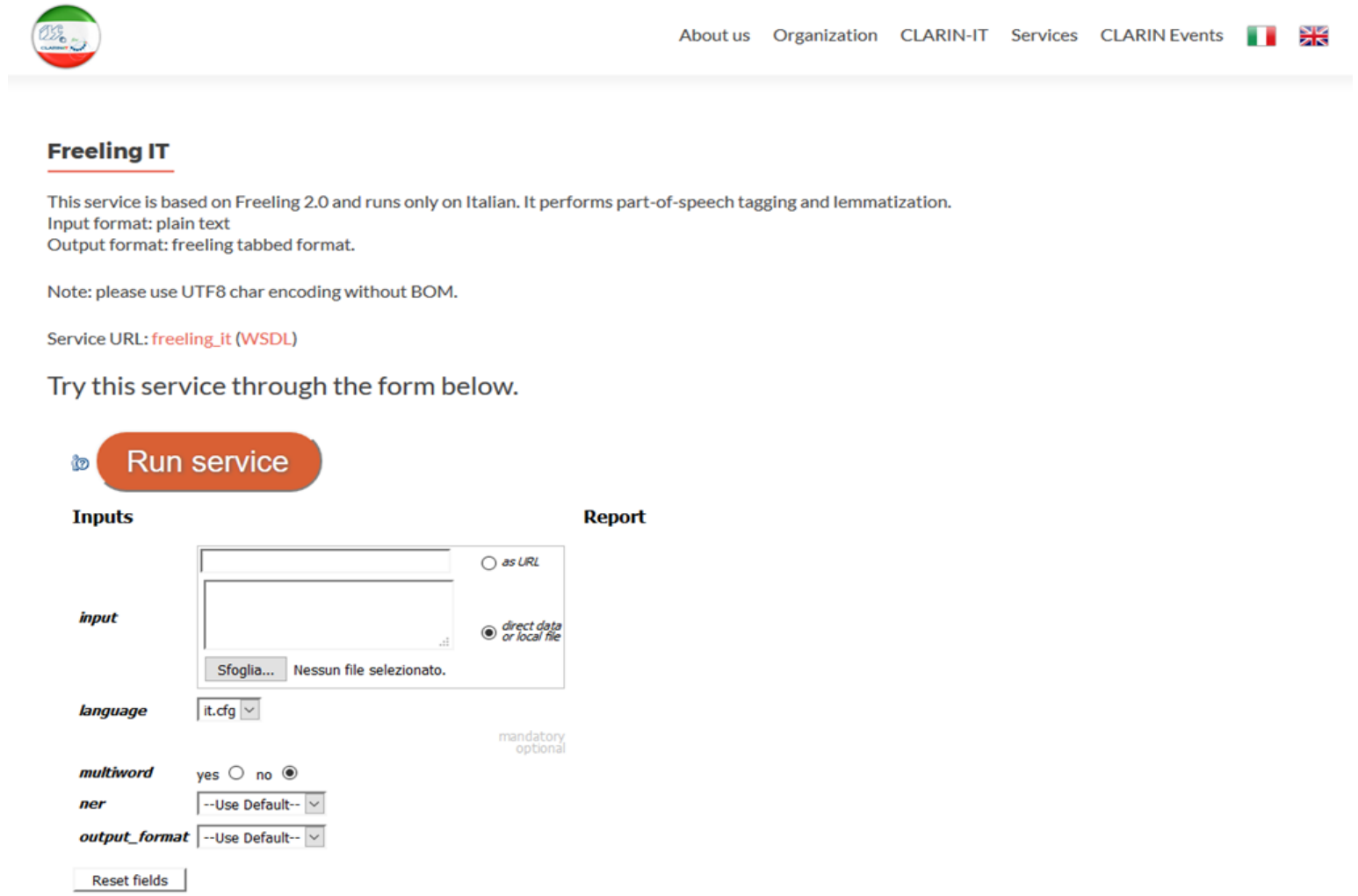
- [1] - edificio o parte di esso in cui si abita;
  (*abitazione* [1], *casa* [1], *dimora* [2], *magione* [1], *ostello* [2], *tetto* [4])

- [2] - edificio con particolari funzioni.
  (*casa* [2])

- [3] - dinastia, casa regnante; l'insieme dei sovrani, appartenenti a una stessa famiglia, che si succedono sul trono
  (*casa* [3], *casa_regnante* [1], *dinastia* [1])
  "la casa di Savoia non ha regnato a lungo sull'Italia"



ILC4CLARIN: strumenti analisi linguistica



Servizi ILC:
altri candidati ... ?



Servizi ILC:
obiettivi a medio e lungo termine



- Migliorare l'esistente:
  - interfacce GUI, usabilità, documentazione, tutorials, etc...;
  - nuove funzionalità;
  - compatibilità Weblicht (REST, TCF);
  - sparql endpoint per risorse RDF / LLOD
  - ...
- Contribuire al FCS

- Creare servizi **NUOVI** in collaborazione con utenti finali (--> SSH, progetti!)

User Involvement: call for proposals



Per eventi di Coinvolgimento Utenti 2017 (Finanziamento da CLARIN-ERIC)

Tre tipi di evento ammissibili:

- **CLARIN Master Classes**;
- **CLARIN Courses at Summer Schools**;
- **CLARIN Training Events**

Per dettagli: https://www.clarin.eu/news/1st-call-clarin-user-involvement-event-proposals-2017

Risorse Linguistiche: bottlenecks

Le risorse

- risiedono presso il produttore
- difficili da reperire
- relegate agli scopi e ambiti scientifici del progetto
- gravate da vincoli





La loro costruzione richiede tempo e denaro: è cruciale

- consentire accesso facile e garantito
- garantire persistenza e fruibilità nel tempo



Risorse Linguistiche: verso un cambio di paradigma

- Nelle scienze dure il processo di data-sharing è dato per scontato

... mentre ...

- tale approccio non rappresenta la norma nelle scienze umane e sociali
- un'attitudine, che impedisce la condivisione dei risultati della ricerca – molti dei quali supportati da fondi pubblici – , rappresenta un FRENO per la scienza

Un cambio di paradigma:

- accesso ai dati, condivisione
- replicabilità dei risultati



Linee strategiche ILC 2017 (dalla riunione personale 13/01)

