

# Dans les coulisses des infrastructures européennes en SHS

Rôle et opportunités pour les acteurs de la  
recherche (ingénieurs et chercheurs)

**Francesca Frontini**

Istituto di Linguistica Computazionale A. Zampolli  
CNR - Pisa



# Qui suis-je?

- Jusqu'à septembre 2020 - MCF en Linguistique Informatique à UPVM
- Membre du laboratoire Praxiling
- Chercheuse à ILC-CNR - Pisa
- Humanité numériques, TAL, linguistique de corpus, e-lexicographie
- CLARIN ERIC ambassador  
CLARIN-IT, CLARIN-FR
- Janvier 2021 - CLARIN Board of Directors



# Plan de la présentation

- INFRAAs en SHS
- Le panorama actuel EOSC
- Exemple concret CLARIN ERIC
- Les opportunités



**Point de  
vue d'une  
chercheuse**



DATA  
INFRASTRUCTURE

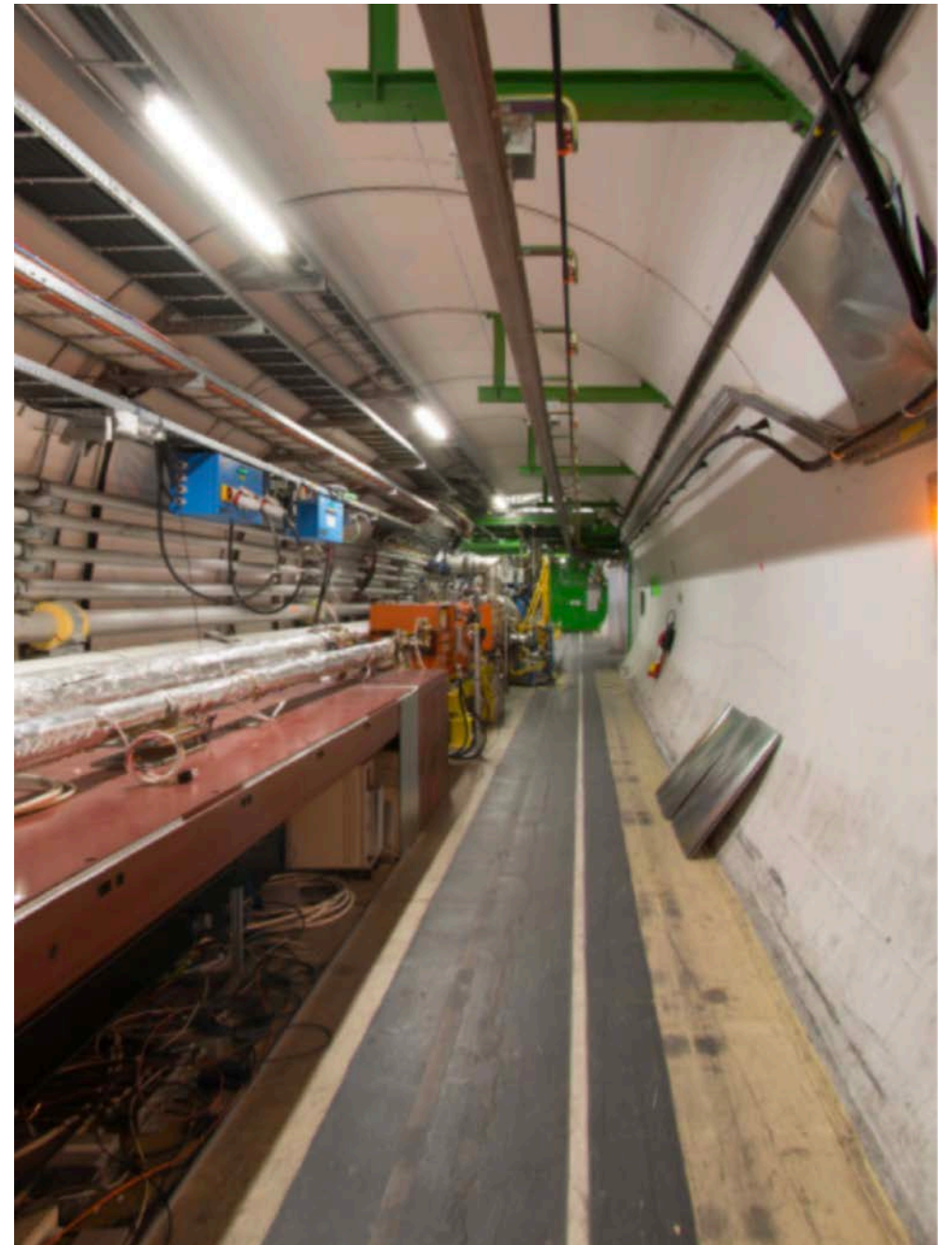


# Infrastructures de recherche

Fournir des ressources et des services aux communautés de scientifiques pour mener des recherches et favoriser l'innovation.

- équipement scientifique majeur ou ensembles d'instruments
- collections, archives ou données scientifiques
- systèmes informatiques et réseaux de communication
- toute autre infrastructure de recherche et d'innovation de nature unique ouverte aux utilisateurs externes

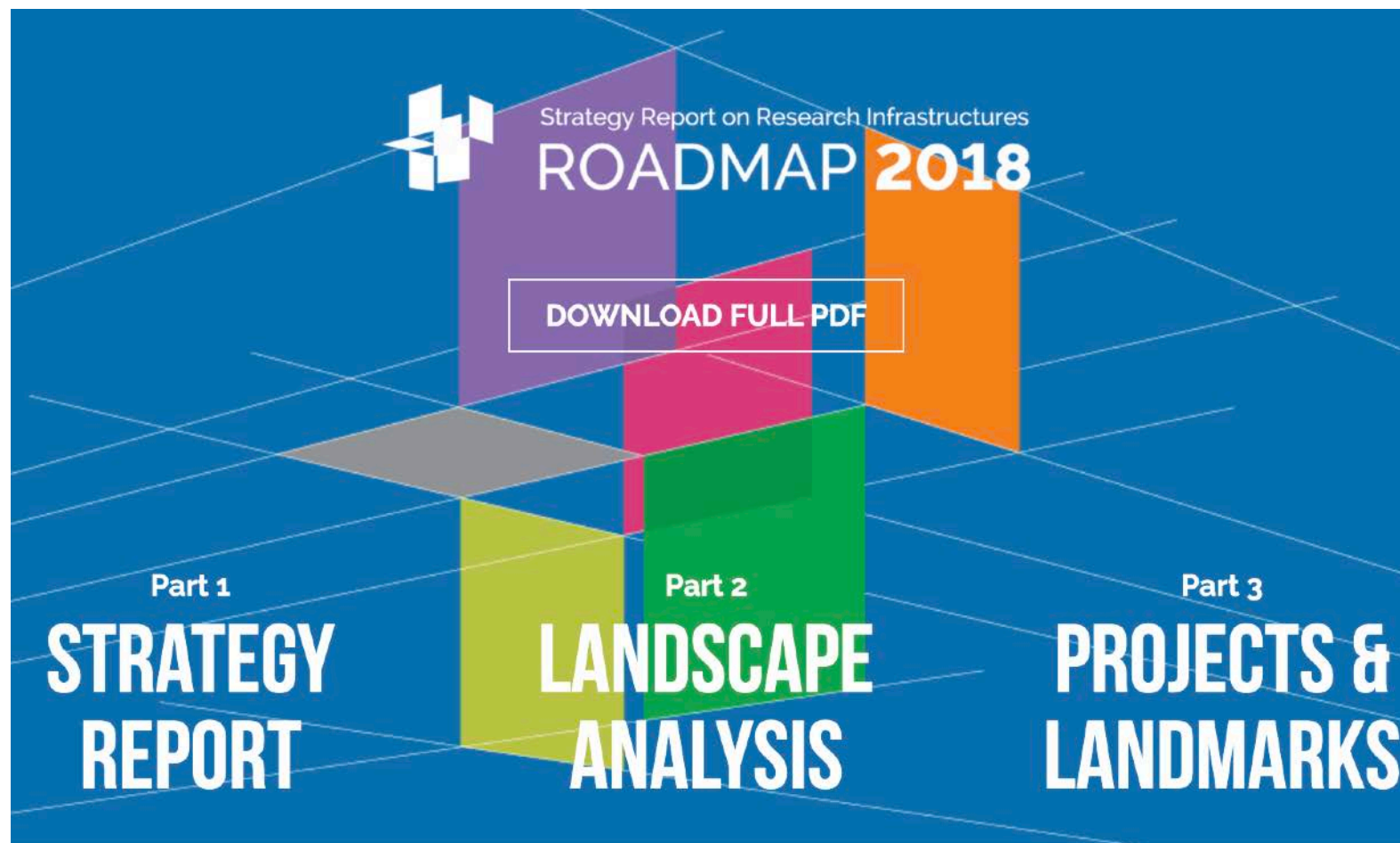
[https://ec.europa.eu/info/research-and-innovation/strategy/european-research-infrastructures\\_en](https://ec.europa.eu/info/research-and-innovation/strategy/european-research-infrastructures_en)



<https://home.cern>

# ESFRI

- [ESFRI.eu](http://ESFRI.eu) - European Strategy Forum on Research Infrastructures (2002-...)



<http://roadmap2018.esfri.eu/>

# Les ERICs



- Phase préparatoire
- Evaluation par ESFRI:
  - Statut legal - ERIC (European Research Infrastructure consortium) & LANDMARK status
  - Financé par l'inscription de chaque nation; soutien par des projets européens
- Single site ou distribuées

[https://ec.europa.eu/info/research-and-innovation/strategy/european-research-infrastructures/eric\\_en](https://ec.europa.eu/info/research-and-innovation/strategy/european-research-infrastructures/eric_en)

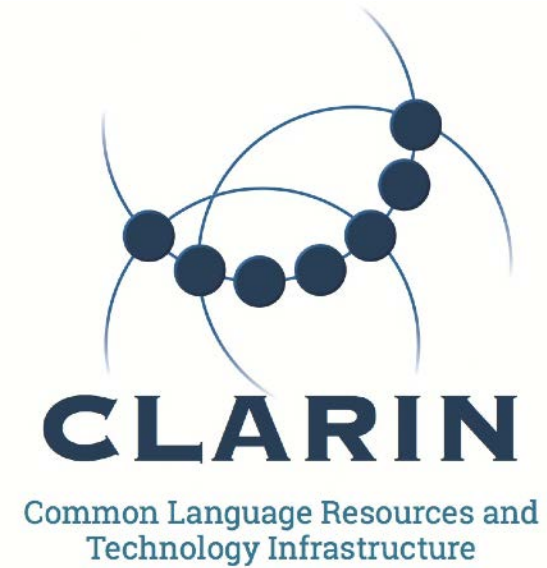
# En SHS

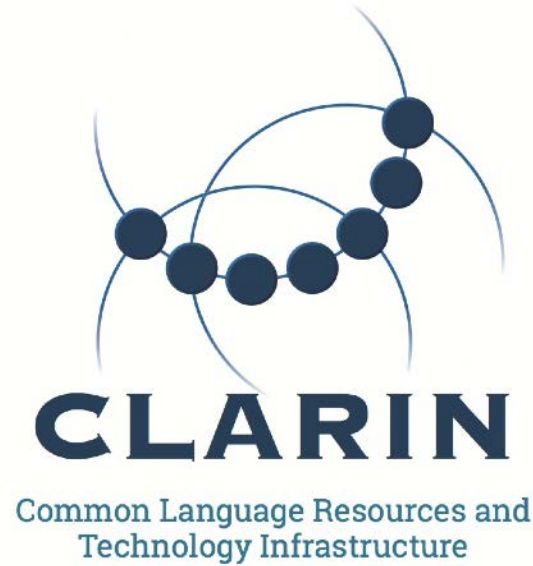
- En SHS; révolution des humanités numériques - data driven, collaborative
- Infrastructures distribuées et virtuelles:
  - données, services, outils, mais aussi compétences, standards, pratiques, savoirs

Research Infrastructures in the Digital Humanities -  
Science Policy Briefing 42 - September 2011  
[https://www.esf.org/fileadmin/user\\_upload/esf/  
RI\\_DigitalHumanities\\_B42\\_2011.pdf](https://www.esf.org/fileadmin/user_upload/esf/RI_DigitalHumanities_B42_2011.pdf)





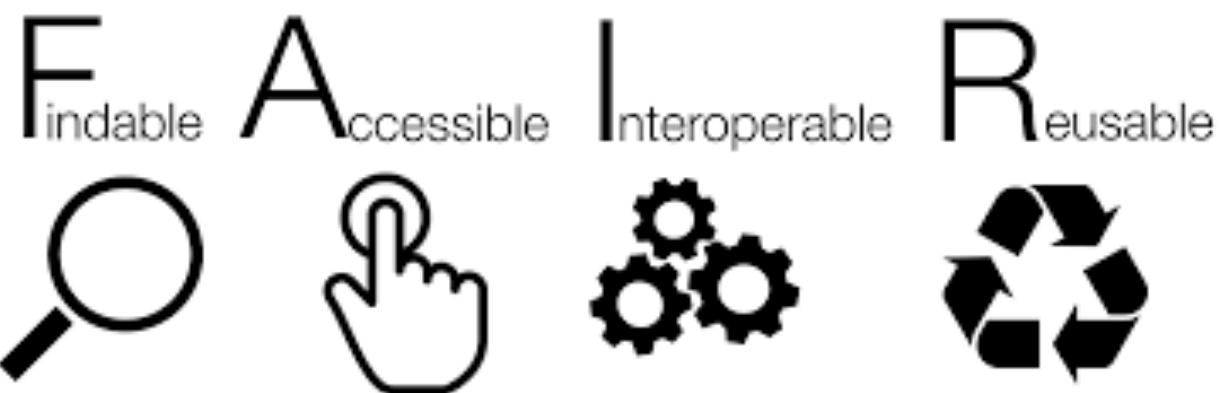






**EUROPEAN OPEN  
SCIENCE CLOUD**

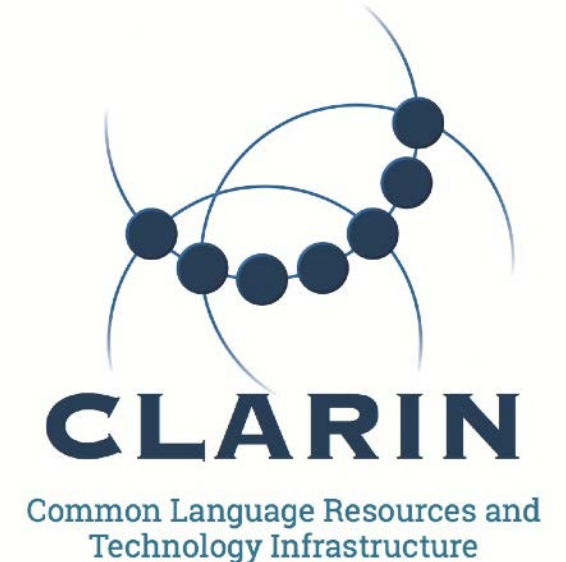
<https://www.eosc-portal.eu/>



<https://www.force11.org/group/fairgroup/fairprinciples>

# CLARIN

- **Common Language Resource Infrastructure for Social Sciences and Humanities**
- Un ERIC, depuis 2012 - ESFRI Landmark depuis 2016 - EOSC ready
- Mission
  - Offrir un accès facile et durable aux chercheurs en sciences humaines et sociales et au-delà
  - aux données linguistiques numériques (sous forme écrite, parlée, vidéo ou multimodale)
  - et à des outils avancés pour les découvrir, les explorer, les exploiter, les annoter, les analyser ou les combiner, où qu'elles se trouvent
- Organisation
  - Federation de centres de donnée
  - Bureau central; consortiums nationaux



# Données et disciplines

- Disciplines:
  - Humanités numériques, linguistique, TAL, histoire, sciences politiques, ...
- Types de ressource
  - Corpus (littéraires, journalistiques, politiques, oraux, ...)
  - Ressources lexicales, outils TAL

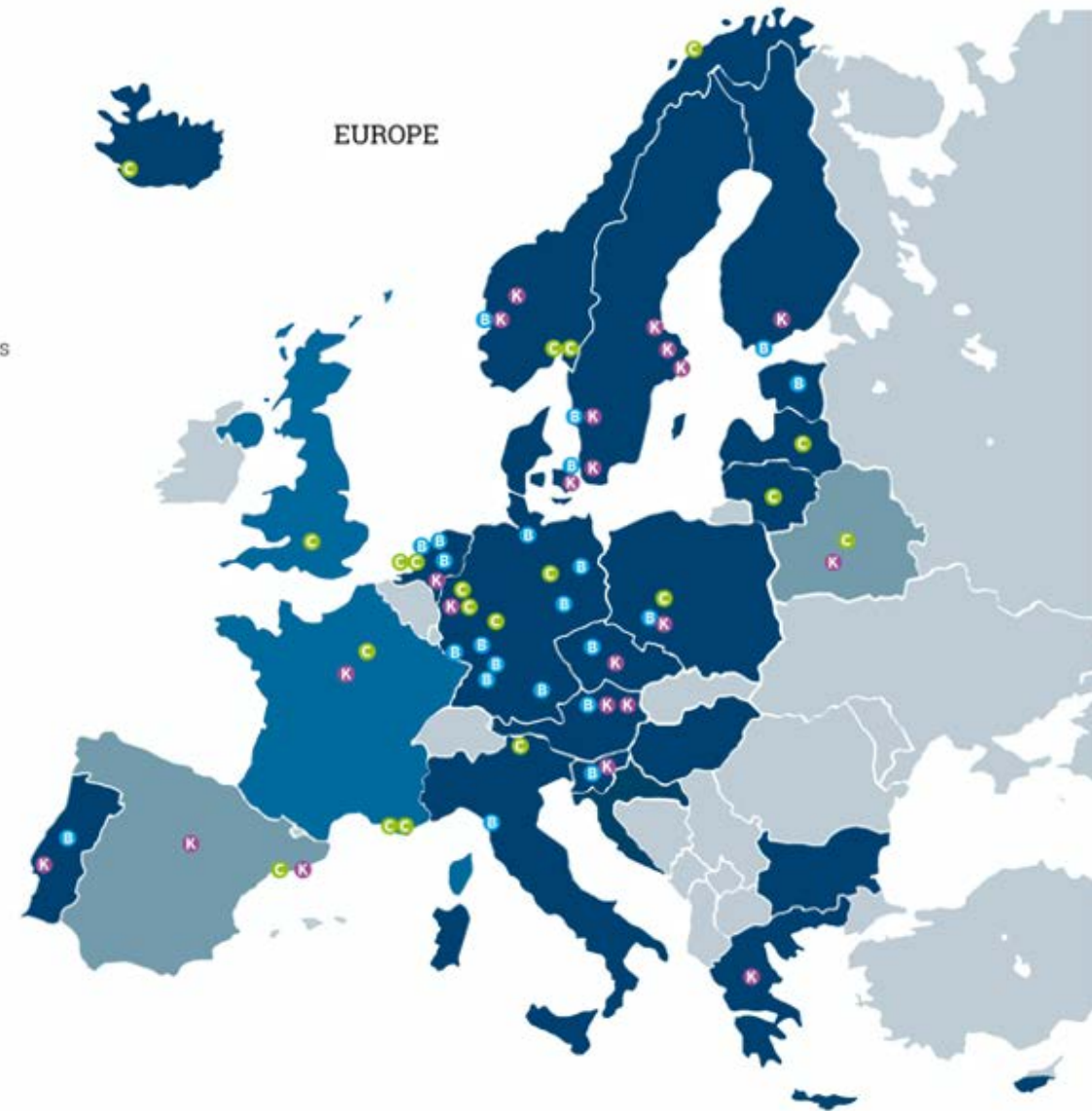


# Les consortiums

- Organisation distribuée
  - Fédération de centres de données (> 50 centres)
  - Bureau central; consortiums nationaux
  - **21 membres: (AT, BG, CY, CZ, DE, DK, EE, FI, GR, HR, HU, IS, IT, LT, LV, NL, NO, PL, PT, SE, SI)**
  - **3 observateurs: FR, UK, ZA**
- Contribution
  - Quota national, services centraux
  - *In kind*, mise à disposition de services nationaux à des chercheurs de tous les pays



- ERIC members
- Observers
- Countries with participating centres
- Centre Providing Data
- Centre Providing Metadata
- Knowledge Centre



# Centres - type

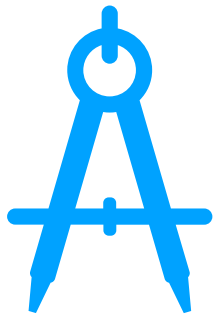
**C**



**Ressources, corpus,  
grammaires formelles et modèles du langage  
ressources lexicales numériques**

....

**B**



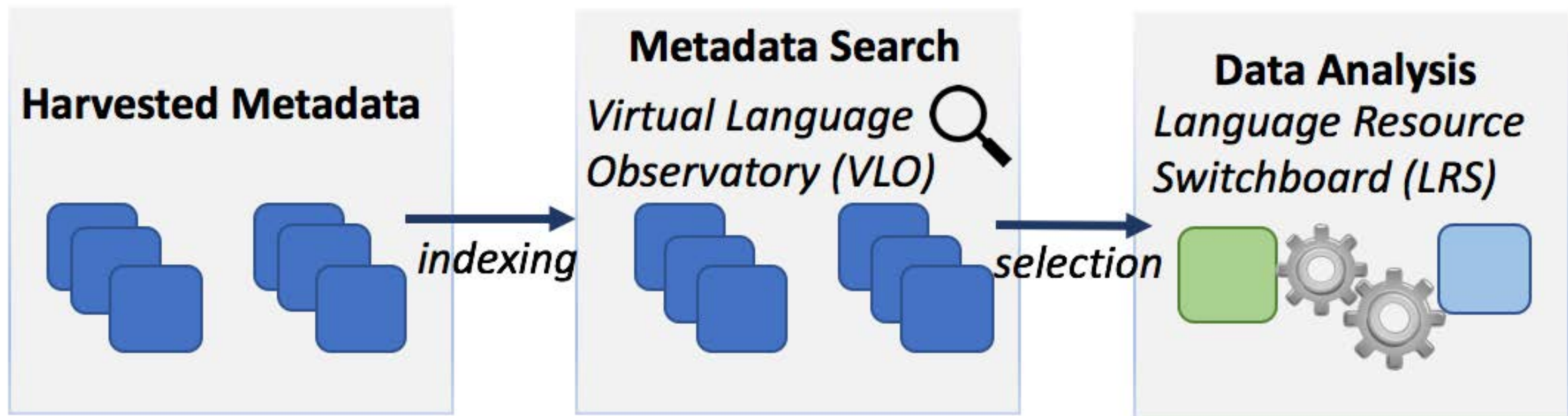
**VREs, interfaces pour l'exploration et l'annotation,  
outils TAL**

**K**



**Compétences**

# clarin.eu



## Outils et Ressources pour un Traitement Optimisé de la LANGue

ORTOLANG est un équipement d'excellence validé dans le cadre des [investissements d'avenir](#). Son but est de proposer une infrastructure en réseau offrant un réservoir de données (corpus, lexiques, dictionnaires, etc.) et d'outils sur la langue et son traitement clairement disponibles et documentés qui :

- permette, au travers d'une véritable mutualisation, à la recherche sur l'analyse, la modélisation et le traitement automatique de notre langue de se hisser au meilleur niveau international;
- facilite l'usage et le transfert des ressources et outils mis en place au sein des laboratoires publics vers les partenaires industriels, en particulier vers les PME qui souvent ne peuvent pas se permettre de développer de telles ressources et outils de traitement de la langue compte tenu de leurs coûts de réalisation;
- valorise le français et les langues de France à travers un partage des connaissances sur notre langue accumulées par les laboratoires publics.




ORTOLANG est un service spécialisé pour la langue, complémentaire de l'offre générale proposée par [Huma-Num](#) (très grande infrastructure de recherche).

La charte d'ORTOLANG définit les modalités d'utilisation et de dépôt de ressources sur la plateforme.

Vous pouvez [consulter la charte](#) ou la télécharger ([fichier au format pdf](#))



 Accueil Corpus Lexiques Terminologies Outils Projets Intégrés Actualités Informations Producteurs

# Corpus 14

Laboratoire producteur :

[Praxiling - UMR 5267 \(Montpellier, FR\)](#)

Soutien institutionnel :

DGLFLF - Délégation générale à la langue française et aux langues de France (Paris, ILF - Institut de linguistique française - FR 2393 (Paris, FR) [↗](#)

Toutes les versions :

 Version 1

 Version 1.1.1

 Version 2

## “ Référence à citer

Texte

BibTeX

Praxiling - UMR 5267 (2019). *Corpus 14* [Corpus]. ORTOLANG (Open Resources and TOols for LANGuage) - [www.ortolang.fr](http://www.ortolang.fr), v2, <https://hdl.handle.net/11403/corpus14/v2>.

## Description

Le projet « Corpus 14 » donne à lire les correspondances de Poilus ordinaires. Il privilégie les écrits peu-lettrés, encore peu exploités par les historiens de la Grande Guerre. Ces documents, mis à disposition par les Archives départementales de l'Ain, l'Ardèche, la Charente-Maritime, l'Hérault, l'Ille-et-Vilaine, la Saône-et-Loire ainsi que par les familles qui en sont les dépositaires, fourniront aux linguistes des ressources nouvelles pour l'histoire de l'écrit peu-lettré.

Le corpus est constitué par la correspondance de trente-sept scripteurs : vingt-cinq hommes et douze



# VLO


poilus



Showing 1 result within selection for poilus French 

Results per page: 10 

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

Language 

French 


Multilingual 

Collection 

Resource type 

## Corpus 14

(Part of [ORTOLANG Repository](#))

 Le projet « Corpus 14 » donne à lire les correspondances de **Poilus** ordinaires. Il privilégie les écrits peu-lettrés, encore peu exploités par les historiens de la Grande Guerre. Ces documents, mis à disposition par les Archives départementales de l'Ain, l'Ardèche, la Charente-Maritime, l'Hérault, l'Ille-et-Vilaine, la ...

French

 [Landing page for this record](#)



# Switchboard

Resource

[submitted\\_text.txt](#)

Mediatype

text/plain

Language

French

## Matching Tools

Search for tool

Group by task

### ▼ Dependency Parsing



Start Tool

UDPipe

### ▼ Distant Reading



Start Tool

Voyant Tools

### ▼ Machine Translation



Start Tool

LINDAT Translation

### ▼ Named Entity Recognition

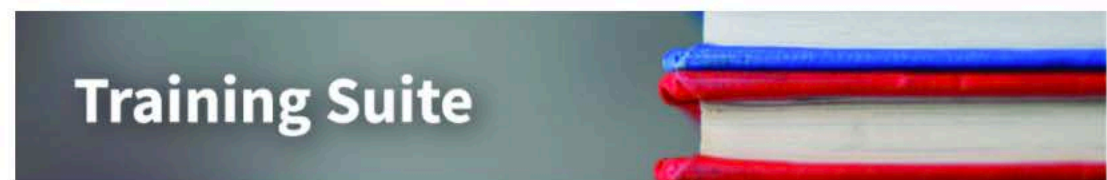
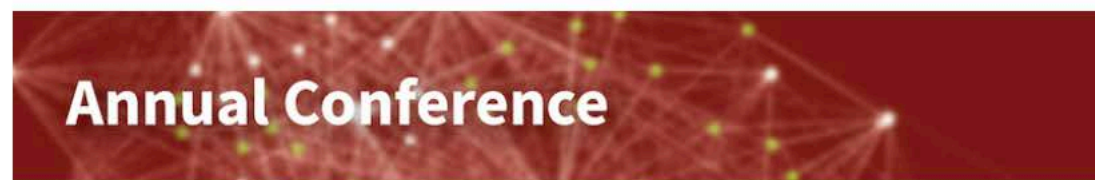
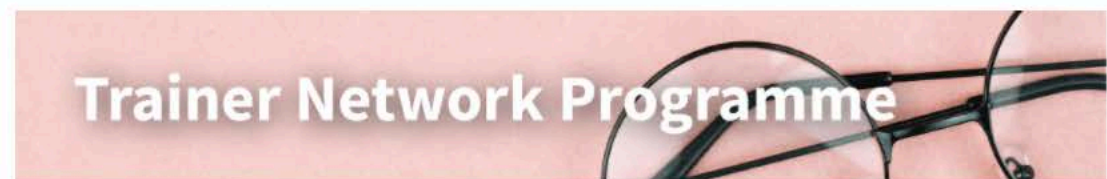
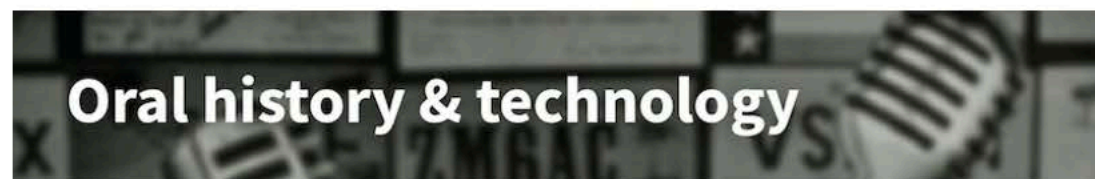
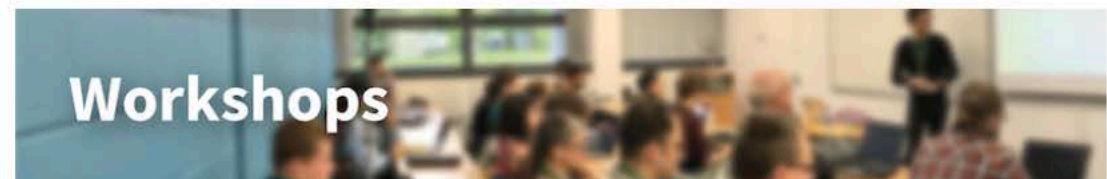


Start Tool

NLP-HUB (multiple NER tools)

# Knowledge sharing

**FIND OUT MORE ABOUT THE MAIN COMPONENTS OF THE CLARIN KNOWLEDGE SHARING INFRASTRUCTURE ...**







FR EN

## CLARIN-FR Consortium

**CLARIN** (Common Language Resources and Technology Infrastructure) is an European infrastructure, structured as an **ERIC** (European Research Infrastructure Consortium), which was created in 2012 to share resources and tools around language at the European level.

France joined **CLARIN** as an observer in 2017. **Huma-Num** coordinates the French participation in CLARIN activities through the consortium **CORLI** a group of experts funded by Huma-Num..

### CLARIN in a nutshell

CLARIN's resources are useful for all disciplines of Social Sciences and Humanities, although the main users are in the fields of linguistics and Natural Language Processing. CLARIN is more specifically aimed at European researchers, but is open to other communities (e.g. research outside Europe, citizens, journalists, etc.).

### French CLARIN centres

Centres-C  
[ORTOLANG](#)  
[Cocoon](#)  
[Phonothèque de la MMSH](#)

Centres-K  
[CORLI Centre K CLARIN](#)

### CLARIN-FR mailing list

To [view the archives](#) of the mailing list  
To [subscribe](#) to the mailing list

### Some language resources

Search tools

[Virtual Language Observatory](#)  
[Federated Content Search](#)

Tools for processing language

[Language Research Switchboard](#)

Collections

[Portail](#)  
[Virtual Collections](#)

Other

[Funding Opportunities](#)

[Knowledge sharing](#)

[Resources for researchers](#)

### Tweets by @CLARINERIC



**CLARIN ERIC**  
@CLARINERIC



Read the interview with Michaela Mahlberg for Tour de CLARIN at [@clarin\\_uk](#) [➔](#)  
[clarin.eu/blog/tour-de-c...](#)



Oct 30, 2020



**CLARIN ERIC**  
@CLARINERIC



Embed

[View on Twitter](#)

<https://clarin-fr.fr/>



WELCOME

CLARIN K CENTRE

CLARIN-FR

PARTICIPATE ▾

ACTIONS ▾

TRAININGS

GUIDELINES ▾

CONTACT ▾

CORLI French CLARIN Knowledge Centre for Corpora, Languages and Interaction



Our team is maintaining a FAQ (Frequently Asked questions) page addressing common concerns in topics related to corpus linguistics. If the FAQ section does not answer your question, you can address your questions to the steering committee of CORLI here:

<https://corli.huma-num.fr/en/contact>.

Our team of experts is doing its best to treat the requests within two working days.



[Other CLARIN K Centres](#)

[The CLARIN website](#)

[CLARIN-FR](#)



# Funding Opportunities

## Notification

As the COVID-19 dynamics are preventing the CLARIN community from conducting face-to-face activities, we warmly invite funding proposals for the preparation of virtual events and other creative formats that could contribute to the continued outreach to communities of use. The normal application guidelines are temporarily suspended. Instead, we ask applicants to convincingly summarize their plans in a format of their choice. If you have any questions about how to prepare your proposal, please do not hesitate to [get in touch](#).

# CLARIN Café

The CLARIN Café is an informal and interactive space where researchers, lecturers and students can meet. The first sessions will be organized by the [CLARIN Ambassadors](#).

Francesca Frontini (Université Paul-Valéry in Montpellier), Maciej Maryl (Polish Academy of Sciences), and Toine Pieters (Utrecht University) will be holding discussions with other experts in the CLARIN community on the topics that we are all thinking about these days, such as:

- how can CLARIN best support researchers who are collecting Covid-19 related datasets
- how to conduct fruitful virtual academic conferences and workshops for CLARIN communities
- how to use CLARIN resources and tools in (virtual) teaching, and
- how beginners from various SSH disciplines with no previous DH experience can incorporate CLARIN resources and tools into their research projects.



< BACK TO START

## DIGITAL HUMANITIES COURSE REGISTRY



The Digital Humanities Course Registry is a curated platform that provides an overview of the growing range of teaching activities in the field of digital humanities worldwide.

The platform is a joint effort of two European research infrastructures: CLARIN-ERIC and DARIAH-EU.



# PARTHENOS





# REGISTER NOW



*Realising the European Open Science Cloud  
Towards a FAIR research data landscape for the  
social sciences, humanities and beyond*

**16-19 November 2020 - EOSC-hub, FREYA and SSHOC**

CAP





# Social sciences and humanities open cloud

Social Sciences & Humanities Open Cloud

Contact us | Join us

zenodo

in

twitter

YouTube

external link

Search



About Training Marketplace Resources News & Events Who Benefits



**SSHOC NEWSLETTER NO. 6**  
**AUGUST PICKINGS**

<https://sshopencloud.eu/>

## Filter by category



Data management



Datasets



Networking



Processing & analysis



Security & operations



Sharing & discovery



Storage



Training & support



# S'informer, participer (1)

- **CLARIN ERIC**

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

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

- <https://www.clarin.eu/content/clarin-cafe>

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



# S'informer, participer (2)

- **CLARIN-FR (Huma-Num)**
  - Coordinateur National **Nicolas Larousse**
  - Responsable UI CLARIN-FR **Adeline Joffres**  
(Responsable de la coopération internationale Huma-Num)
  - <https://www.clarin-fr.fr/>
  - <https://listes.huma-num.fr/sympa/subscribe/clarin-fr>
- **SSHOC project - Training & events (Webinars)**
  - <https://sshopencloud.eu/>

## Workshops & Webinars

In 2020 and 2021, SSHOC will run regular **webinars** and **workshops** on themes including:

- Data Protection and the GDPR
  - Workshop: TBA
  - Webinar: [Putting Data Protection into Practice: GDPR and the DARIAH ELDAH Consent Form Wizard](#)
- Data Stewardship and RDM in theory and practice
  - Workshop: [Caring for Sharing – Data Management and FAIRness of Migration Data](#)
  - Webinar: [Tools and resources for FAIR data](#)
  - Webinar: [Introducing the newly launched EMM Survey Registry](#)
- Data Citation principles and practice
  - Workshop: TBA
  - Webinar: TBA
- Data Science for the Social Science and Humanities
  - Workshop: [Digital Humanities 2019 Workshop: The Case Of Interview Data](#)
  - Webinar: [CLARIN Hands-on Tutorial on Transcribing Interview Data](#)
  - Webinar: [Sharing Datasets of Pathological Speech](#)
- Data Science for Heritage Science
  - Workshop: TBA
  - Webinar: [Use and Re-use of Scientific Data in Archaeology and Heritage](#)
- Text Mining for the Social Sciences and Digital Humanities
  - Workshop: [Using Corpora for Implementing Validation. Workflows that combine quantity and quality](#)
  - Webinar: [Quanlify with Ease: Combining quantitative and qualitative corpus analysis](#)

# Le role des ingénieurs

## Retour d'investissement

- Connaitre le panorama des services et infrastructures
- Pousser l'utilisation dans les laboratoires, projets, etc.

## Retour d'experience

- Faire remonter aux infrastructures les besoins
- Promouvoir la participation active dans la co-construction des services

# Les ingénieurs dans les ERIC

- Les ERICS emploient directement un nombre d'ingénieurs
  - **Dieter Van Uytvanck** est le directeur technique CLARIN ERIC
  - Task forces et comités (standard, métadonnées, content search)
- Dans les pays membres, opportunités de collaborer aux groupes de travail des consortiums nationaux



# EURISE Network

Online expert meeting series 2020

Date: 28 October 2020 - 15:00 to 09 December 2020 - 16:00

**European Research Infrastructure Software Engineers' Network (EURISE)**

<https://eurise-network.github.io/>



**Merci de votre attention**

*francesca.frontini@ilc.cnr.it*