

# How Not to Spill Coffee on Your Tapes. Best Practices for Preserving Oral Archives

**CLARIN ERIC and SSHOC Café**  
Wednesday the 24th of February 2021



# Organisers

**Silvia Calamai**, Associate Professor of linguistics and sociolinguistics at the Department of Education, Human Sciences and Intercultural Communication at the University of Siena and a member of CLARIN-IT

**Véronique Ginouvès**, Research Engineer CNRS, in charge of the CLARIN-FR centre *Phonothèque MMSH - Maison méditerranéenne des sciences de l'homme* in Aix-en-Provence.

<http://phonothèque.mmsh.huma-num.fr/>

<http://www.clarin-it.it/it/content/archivio-vivo>

# Plan

- Intro:
  - Francesca Frontini (ILC-CNR & CLARIN ERIC) - *CLARIN 101*
  - Monica Monachini (ILC-CNR & CLARIN-IT) - *The SSHOC project*
- The café:
  - Silvia Calamai - *The experience of the Italian "Vademecum"*
  - Véronique Ginouvès - *Sonorités* and the Michel Seurat sound archives, with Claire Cialone-Gregoire
- Discussion
- Closing remarks

The event is recorded for further dissemination purposes.

**Questions and comments?** Put them in the chat box.

# CLARIN 101

<https://www.clarin.eu/content/clarin-in-a-nutshell>

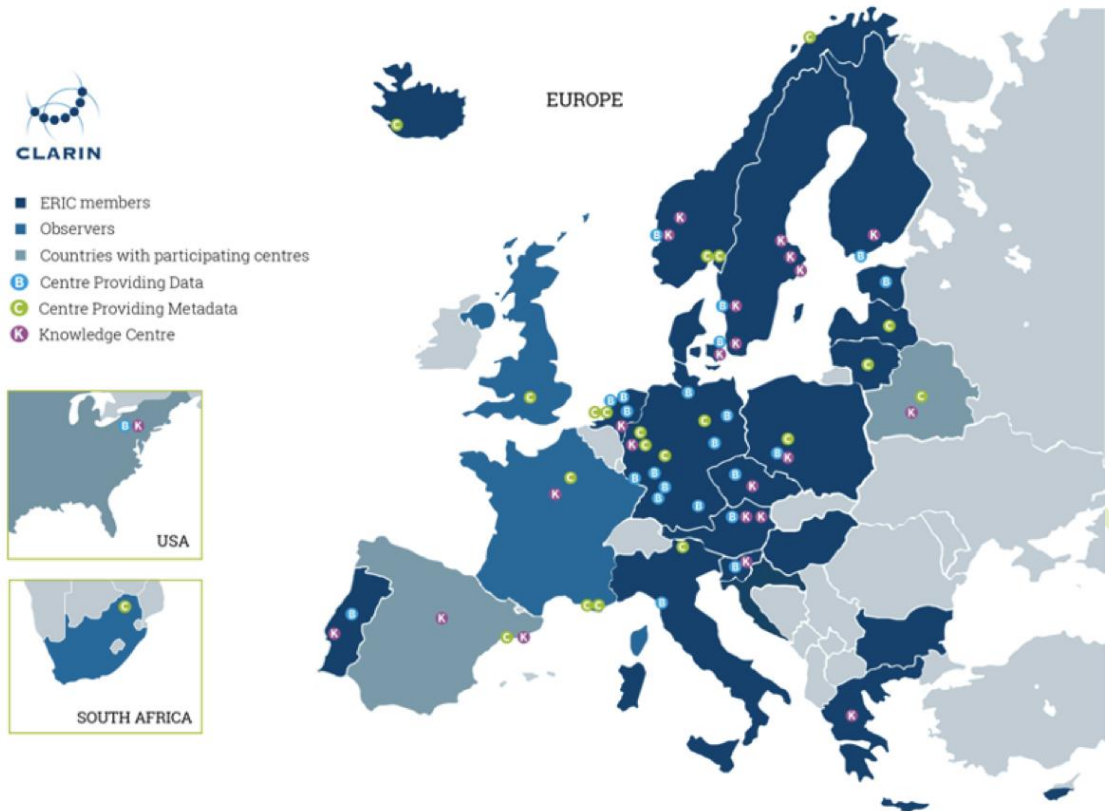


# CLARIN in eight bullets

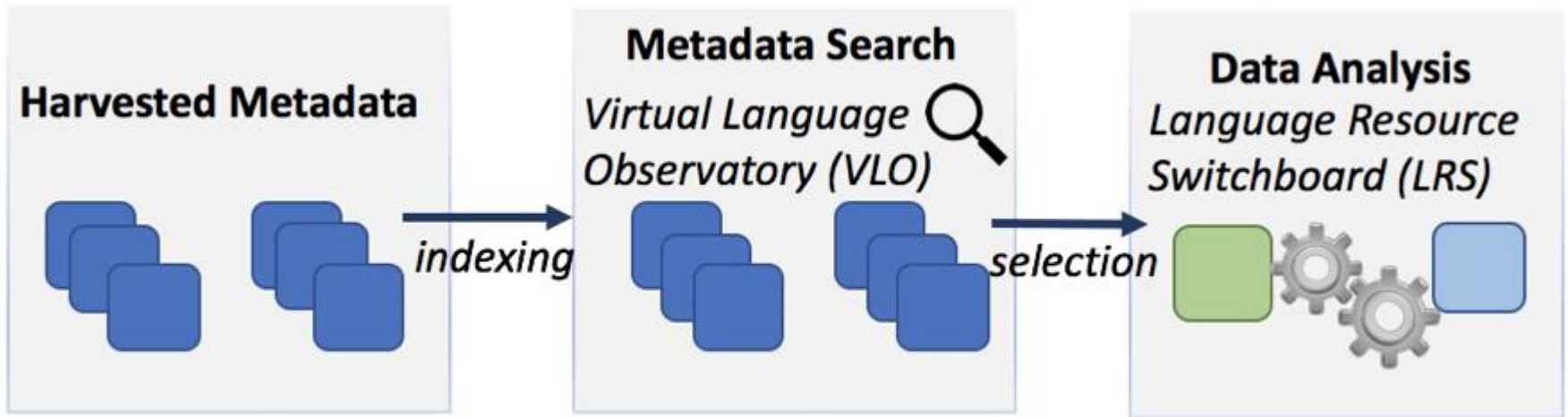
- **CLARIN** is the Common Language Resources and Technology Infrastructure
- **ESFRI** ERIC status since 2012, Landmark since 2016
- that provides easy and sustainable access for scholars in the **humanities and social sciences** and beyond
- to **digital language data** (in written, spoken, video or multimodal form)
- and **advanced tools** to discover, explore, exploit, annotate, analyse or combine them, wherever they are located
- through a **single sign-on** environment
- that serves as an ecosystem for **knowledge sharing**
- and: ready for **integration in EOSC**

# CLARIN today

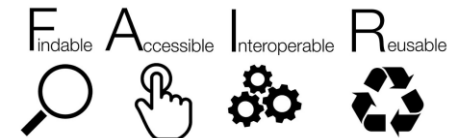
- **> 50 centres**
- **21 members:** (AT, BG, CY, CZ, DE, DK, EE, FI, GR, HR, HU, IS, IT, LT, LV, NL, NO, PL, PT, SE, SI)
- 3 observers: FR, UK, ZA



# The technical Infrastructure



<https://vlo.clarin.eu/>  
<https://switchboard.clarin.eu/>



# The Knowledge Infrastructure



Knowledge centres



VideoLectures



Funding for User Involvement events



Workshops



Digital Humanities Course Registry



CLARIN Mobility Grants



Trainer Network Programme



Training Suite

<https://www.clarin.eu/content/clarin-for-researchers>

<https://www.clarin.eu/content/knowledge-sharing>



# CLARIN for Oral History

<https://www.clarin.eu/oral-history>

Project:



Horizon 2020  
European Union Funding  
for Research & Innovation

**Type of action & funding:**  
Research and Innovation action  
(INFRAEOSC-04-2018)

**Partners: 45**

(20 beneficiaries + 25 LTPs)

SSH ESFRI Landmarks and Projects  
& international SSH data infrastructures

**Project budget:**

€ 14,455,594.08

**Duration: 40 months**

(January 2019 – 30 April 2022)

**Project website:**  
[www.SSHopencloud.eu](http://www.SSHopencloud.eu)



**Objectives:**

- creating the social sciences and humanities (SSH) part of European Open Science Cloud (EOSC)
- maximising re-use through Open Science and FAIR principles (standards, common catalogue, access control, semantic techniques, training)
- interconnecting existing and new infrastructures (clustered cloud infrastructures)
- establishing appropriate governance model for SSH-EOSC

# SSHOC - Expected Impact



The Social Sciences and Humanities are seamlessly integrated in the European Open Science Cloud



Availability of an EU-wide, easy-to-use SSH Open Marketplace, where tools and data are openly accessible



EU-wide availability of high quality "cloud ready" SSH tools and high quality SSH data



EU-wide availability of trusted and secure access mechanisms for SSH data, conforming to EU legal requirements



State of the art Research Infrastructure in several pilot domains advanced through dedicated SSH data pilots cluster projects



Maximising reuse through Open Science and FAIR principles (standards, common catalogue, access control, semantic techniques, training)

# This Café and SSHOC (1/2)

This Café is a continuation of events dealing with **spoken data**:

- Working with interview data - July 2019

*how audio material can be digitised, converted by ASR and processed with NLP tools*

- Tutorial on Transcribing Interview Data - March 2020

*workflow of machine-assisted interview transcription process*

- Sharing Datasets of Pathological Speech - Nov 2020

*speech corpora: costly to obtain, hard to share*

... but also dealing **data management and protection**

- Tools and Resources for FAIR Data - May 2020

*help researchers prepare their data in adherence with FAIR guidelines*

- Putting Data Protection into Practice - Oct 2020

*How does the GDPR impact the work of SSH researchers? How RI can support them in these challenges*



# This Café and SSHOC (2/2)

This Café is a part of the SSHOC

## **awareness raising events**

- to support various research communities in the Social Sciences and Humanities as well as other interested individuals
- to promote the outcomes of the SSHOC project
- to meet the needs for new skills

Check the awareness raising & training page on the SSHOC website:

<https://sshopencloud.eu/awareness-raising-events>

<https://sshopencloud.eu/training>

# **The experience of the Italian “Vademecum”**



*Who's afraid of  
oral archives?*







# The starting point

Oral sources as fragile objects

Fragmentation of projects, initiatives, repositories

Absence of shared guidelines

Different types of users

# Chronicle

**XV AISV Congress, Arezzo 16.02.2019: Round Table on Oral Archives**

**20.05.2019, SABBOS, Firenze: 1st working meeting**



# The participants: people and institutions

Coordinated by: Silvia Calamai (Università di Siena – AISV), Alessandro Casellato (Università Ca' Foscari Venezia – AISO), Maria Francesca Stamuli (SABTOS)

## *Conservation of oral archives:*

Sergio Canazza (CSC, Università di Padova), Alessandro Casellato (AISO, Università Ca' Foscari Venezia), Giovanni Contini (AISO), Luciano D'Aleo (ICBSA), Duccio Piccardi (AISV, Università di Siena), Niccolò Pretto (CSC, Università di Padova), Elisa Salvalaggio (Istituto Parri, Istituto piemontese per la storia della Resistenza e della società contemporanea).

## *Description of oral archives:*

Piero Cavallari (ICBSA), Antonella Fischetti (ICBSA), Elena Musumeci (ICCD), Diego Robotti (Soprintendenza Archivistica e Bibliografica del Piemonte e della Val d'Aosta), Alessandro Russo (CSC, Università di Padova), Carla Scognamiglio (ICCU), Maria Francesca Stamuli (Soprintendenza Archivistica e Bibliografica della Toscana),

## *Use and reuse of oral archives:*

Silvia Calamai (AISV, Università degli Studi di Siena, CLARIN-IT), Silvia Filippin (DGA), Sabina Magrini (ICBSA), Monica Monachini (ILC "A. Zampolli", CNR e CLARIN-IT), Giulia Piperno (ICBSA), Niccolò Pretto (ILC– CNR), Micaela Procaccia (ANAI), Sara Zanisi, Fondazione ISEC-Rete Parri.

# Non di sola carta. Prendersi cura degli archivi orali

**Martedì 27 ottobre 2020 - ore 10-13**

**diretta Facebook:** <https://www.facebook.com/ICBSAItalia/videos/357512535359383/>

10.00 - Apertura dei lavori e saluti di Carlo Birrozzi, Direttore ICBSA

10.15-11.15 - **La digitalizzazione degli archivi orali in Italia: un intervento urgente e indifferibile**

Introduzione di Alessandro Casellato (AISO - Università Ca' Foscari Venezia)

Interventi di Paolo Masini (Presidente Museo dell'Emigrazione Italiana), Laura Moro (Dirigente Generale Istituto Centrale per la Digitalizzazione del Patrimonio Culturale – Digital Library), Valdo Spini (Presidente Associazione delle Istituzioni di Cultura Italiane)

Intermezzo e pausa: **in ascolto**

11.30-12.00 - **Perché un “Vademecum per il trattamento delle fonti orali”**

Interventi di Silvia Calamai (AISV - Università di Siena), Maria Francesca Stamuli (Soprintendenza Archivistica e Bibliografica della Toscana)

12 - **Interventi e domande dei partecipanti e del pubblico collegato**

13 - Chiusura dei lavori

 Documentazione utile:

- [VADEMECUM 1.0\\_1. PRODUZIONE E DESCRIZIONE.pdf](#)
- [VADEMECUM 1.0\\_2. CONSERVAZIONE.pdf](#)
- [VADEMECUM 1.0\\_3. USO E RIUSO.pdf](#)
- [VADEMECUM 1.0\\_BIBLIOGRAFIA E SITOGRAFIA.pdf](#)

Fino al 15 gennaio 2021 il Vademecum resterà aperto a osservazioni e proposte che potranno essere inviate rispondendo al questionario reperibile al seguente indirizzo: <https://forms.gle/U48rUuk2zDrrig5R7>

Sarà successivamente approvato e pubblicato nella versione definitiva nel corso del 2021.

# The structure of the Vademecum

<http://www.clarin-it.it/it/content/vademecum-il-trattamento-delle-fonti-orali>

Il Vademecum si compone di tre documenti e relative appendici.

Documenti:

- *Produzione e descrizione delle fonti orali*: come impostare una scheda di rilevazione delle fonti orali e come descrivere un archivio orale.
- *Conservazione degli archivi orali*: come procedere alla salvaguardia delle fonti orali registrate nei decenni passati, in considerazione della loro peculiare fragilità.
- *Valorizzazione, uso e riuso delle fonti orali*: il quadro normativo da tenere presente prima di iniziare una ricerca con fonti orali e indicazioni su dove depositare gli archivi orali e come renderli accessibili ad altri.

Appendici:

- Elenco dei centri specializzati nella digitalizzazione di documenti sonori
- Elenco dei centri di conservazione degli archivi orali in Italia
- Modelli di riferimento per informativa e consenso ai sensi del GDPR
- Modelli di atti di donazione e legato testamentario
- Modelli di contratti di deposito e comodato per la cessione di un archivio orale
- Esempi di riuso di archivi orali del passato
- Istruzioni per acquisire documenti audio e audiovisivo
- Modello di scheda di rilevazione
- Bibliografia e Sitografia

# Public review (Nov. 2020-Feb.2021)

18 answers to an online questionnaire

5 reviews via e-mail

who answered?

scholars, PhD students, archivists, librarians,  
documentalists

degree of satisfaction: 3.05 (scale 1-4)

# What's next

Final version of the Vademecum (summer 2021)

Public presentation (27.10.2021 World Day for Audiovisual Heritage)

Short version and English version

Tutorials and dissemination actions among students, PhD students, librarians...

# **The *Sonorités* Issues**



# AFAS - French Association of Sound, Oral and Audiovisual Archives (created in 1979)

The AFAS journal, *Sonorités*, focuses on the oral source used in humanities and social sciences, the issues and the scientific and cultural assets of the oral testimony.



2 issues in 2020



SONORITÉS

# Sonorités, a journal in openaccess

<https://journals.openedition.org/afas>



[Ai rescontra ma mia](#) (song from Provence) recorded on a cylinder in 1900 at the Universal Exhibition in Paris, Nanterre archives, [listening on Telemeta data base \(Nanterre\)](#).

Issue directed by Marie-Barbara Le Gonidec, ethnomusicologist

Sound archive [3268](#), Interview by [Michel Seurat](#) with a witness of the Hama massacre, Syria, 1982. [Listening on Ganoub](#), the sound archives of the MMSH ([available here](#))

Issue directed by Claire Cialone-Gregoire, archivist

HORS SÉRIE N° 2  
CARNÉTOUR LOCALISME  
Fonction, Mémoire, Terrain  
muséologique et sonore

Éditorial  
Michel Seurat / Les archives sonores  
de la collection Lyndberg  
Portrait de Michel Seurat  
Polémiques du fait, Michel Seurat et  
de son territoire  
Liste des contributeurs  
Comptes rendus / Réponses de terrain par  
Michel Seurat  
Comptes rendus / Les archives sonores de  
Michel Seurat  
Comptes rendus / Réponses de terrain par  
Michel Seurat

Édition et diffusion  
Muséum de Nanterre  
Muséum de Paris  
AFAS

SONORITÉS

# Getting involved in CLARIN

- Join our NewsFlash
  - <https://www.clarin.eu/content/newsflash>
- Check out our events
  - <https://www.clarin.eu/events>
- Open calls
  - <https://www.clarin.eu/content/funding-opportunities>
- Follow us on Twitter @CLARINERIC
- And stay tuned for the next cafés
  - <https://www.clarin.eu/content/clarin-cafe>
  - **#clarincafe**

# Getting involved in SSHOC

Join our community



<https://www.sshopencloud.eu>



[info@sshopencloud.eu](mailto:info@sshopencloud.eu)



[@SSHOpenCloud](https://twitter.com/SSHOpenCloud)



[/in/sshopencloud](https://www.linkedin.com/company/sshopencloud)

# See you at the next café

**CLARIN Café on the rights of data subjects in language resources,** by the CLARIN Legal and Ethical Issues Committee

30 March 2021 - 14.00-16.00 CET



# Additional links

Véronique Ginouvès & Isabelle Gras, dir. *Guide des bonnes pratiques éthiques et juridiques*

<https://presses-universitaires.univ-amu.fr/diffusion-numerique-donnees-shs>

Claire Marcadé, Bernard Guinard, Stéphanie Coulais, Yvon Davy, Eric Desgrugillers, et al.. *Patrimoine culturel immatériel. Traitement documentaire des archives sonores inédites*. Guide des bonnes pratiques.. 2014, pp.82. (<[halshs-01065125](https://halshs.archives-ouvertes.fr/halshs-01065125)>)

Italian Vademecum for the Treatment of Oral Sources

<http://www.clarin-it.it/en/content/vademecum-treatment-oral-sources>