Spanish CLARIN K-centre:
Tecnologías del Lenguaje para Humanidades y Ciencias Sociales (y más allá)

www.clarin-es.org
www.clarin.eu

Mikel Iruskieta
Grupo Ixa
HiTZ Basque center for Language Technology
UPV/EHU

CLARIN

eman ta zabal zazu
Universidad del País Vasco
Euskal Herriko Unibertsitatea

HiTZ
Hizkuntza Teknologiako Zentroa
Basque Center for Language Technology

CLARIN CENTRE K INTELE DARIAH-EU
Infraestructura de Tecnologías del Lenguaje

CLARIN

ESFRI European Strategy Forum on Research Infrastructures

ESFRI PL Projects Landmarks ROADMAP 2016

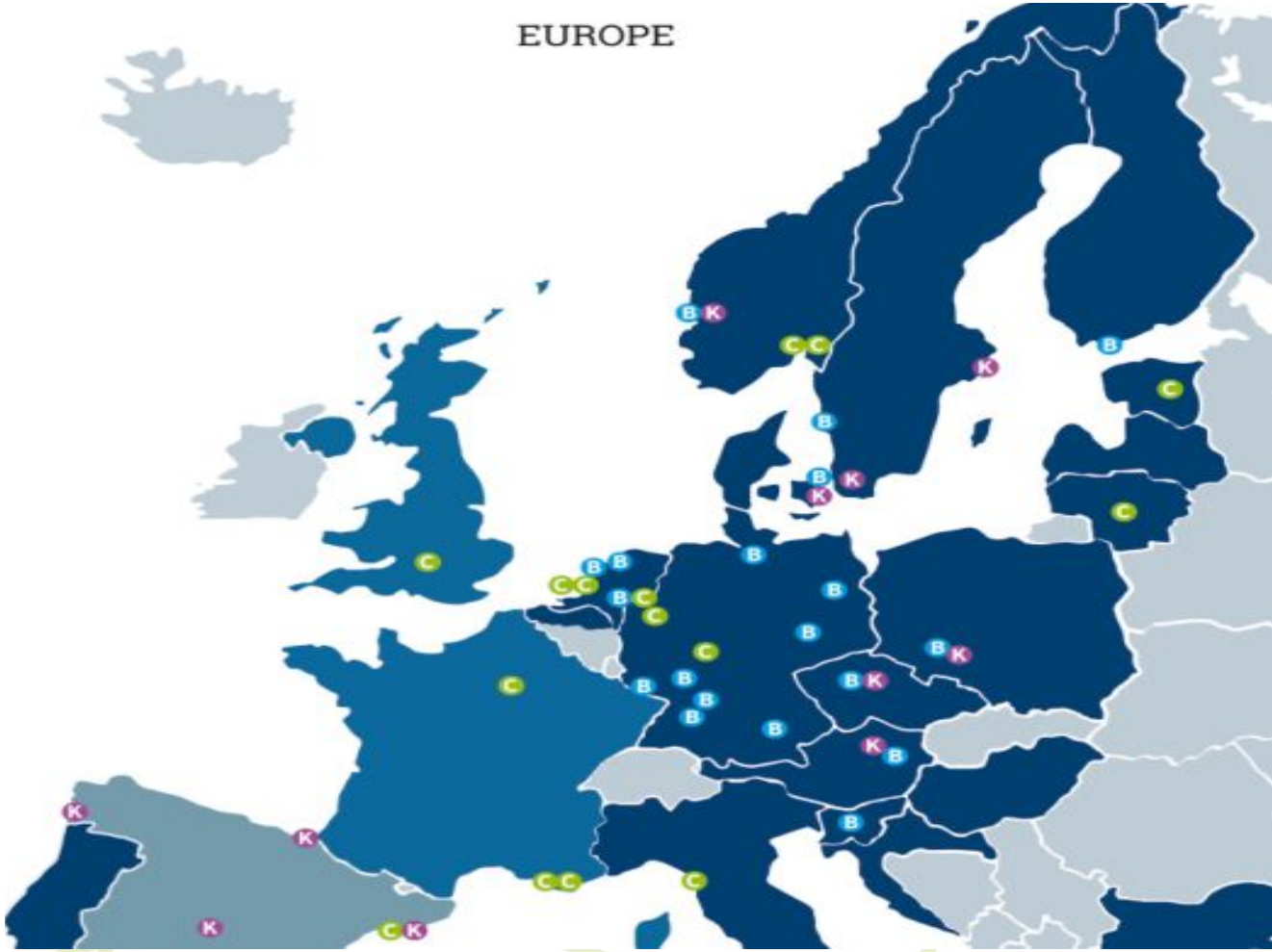
- **C**ommon **LA**nguage **R**esources and Technology **INfrastructure**
  - ESFRI ERIC (2012) y ESFRI Landmark (2016)

- El objetivo es dar a**cceso confederado a los datos y herramientas** para usar tecnología del lenguaje en los estudios de todas las lenguas europeas **desde una única web** y así ayudar a los investigadores de las Humanidades y Ciencias Sociales (y más allá)
  - Datos lingüísticos digitales (textos, voz, multimodales)
  - Herramientas avanzadas: describir, analizar y comparar

Strategy Report on Research Infrastructures ROADMAP 2018

CLARIN

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



Núria Bel
UPF

Mikel Iruskieta
UPV/EHU

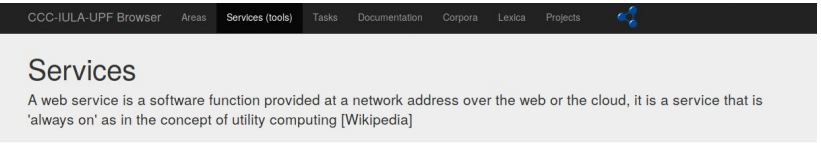
Salvador Ros
UNED

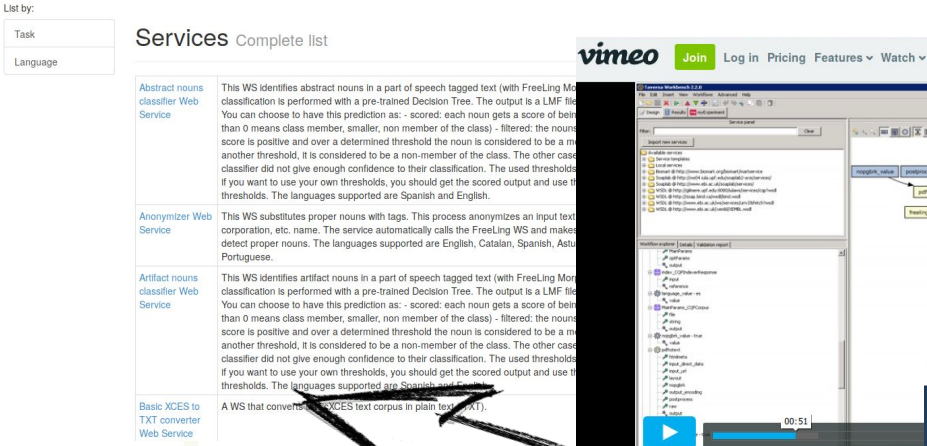
Xavier Gómez Guinovart UVIGO

CLARIN CENTRE K

Implemetación de CLARIN en España

Herramientas y difusión realizado por CCC-IULA-UPF (primer centro CLARIN)



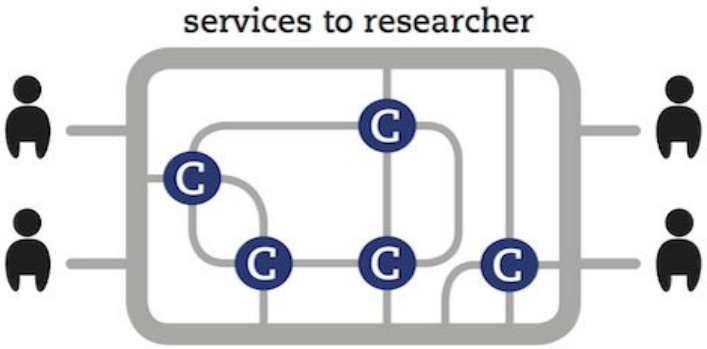


¿?

Objetivos de CLARIN-K

- Potenciar el **acceso a instrumentos tecnológicos** en la investigación en ámbitos de las Humanidades y Ciencias Sociales en español y lenguas de España.

- CLARIN sea la infraestructura que capacite a las HH y CCSS para obtener beneficios de:
  - **acceso** unificado, compartido y colaborativo a recursos digitales distribuidos, y
  - **análisis de grandes cantidades** de datos para que pueda haber un cambio de magnitudes, acortamiento de los ciclos de investigación, etc.



CLARIN-K Centro Español: evaluación interna (2017)

1º centro de competencias evaluado por CLARIN

- Participación en comisiones europeas
  - KSI (Knowledge Sharing Infrastructure Committee)
  - SAB (Scientific Advisory Board)
- Página web (abril 2016): www.clarin-es.org
  - Contacto y respuesta en menos de 2 días
  - Descripción de los grupos y algunas herramientas
- Herramientas de fácil uso para todas las lenguas
  - Contawords, Analhitza, DContado
  - Asesoramiento a grupos hispanoamericanos sobre las herramientas
  - Difusión en HDH
- Asesoramiento en varias tesis

Para preparar la participación de España en la infraestructura europea como CLARIN-ES se firmarón 14 convenios en 2009: **necesidad de actualizar y aumentar para un re-evaluación positiva** para entrar en CLARIN

CLARIN-K Centro Español: evaluaciones internas (3)

- Evaluaciones.
  - 2015-2017
  - 2017-2019
  - 2019-2020 ---> anualmente
- Servicios a 4/6 investigadores por año
- Asesoría a organismos de la Administración
- Implementación de algunas herramientas y métodos
- Publicaciones y conferencias :
  - Bel, N. Gonzalez-Blanco, E. Iruskieta, M. (2016).
    - CLARIN Centro-K-español. Procesamiento del Lenguaje Natural 57: 151-154. ISSN: 1135-5948.
  - Iruskieta, M. Bel, N. (2017).
    - CLARIN-K Centre Spain: una infraestructura orientada usuario. LINHD-UNED. Escuela de Verano HD.

Servicios de CLARIN-K

publicado en el Tour de CLARIN

- Tesis doctoral del Basque Center on Cognition, Brain and Language (BCBL)
- Investiga la adquisición del lenguaje a nivel cognitivo y neuronal y cuánto tiempo de exposición se necesita para cada lengua en un contexto bilingüe, usando:
  - ANALHITZA
  - https://switchboard.clarin.eu

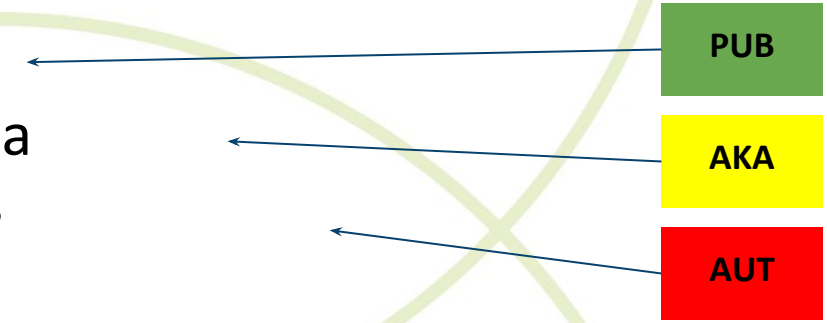
https://www.clarin.eu/blog/tour-de-clarin-interview-jose-perez-navarro

CLARIN K-centre desde el 2016 ha pasado 3 evaluaciones, ha creado herramientas de fácil uso y ha ofertado de 4 a 6 servicios anuales.

CLARIN es mucho más que un proyecto

- Acceso confederado a todos los recursos y datos desde una única web
- **Estándares**
- **Protocolos** comunes
- Ayuda para el cambio de paradigma
- Diseño de los recursos **estratégicos**

- Corpus abiertos y **púb**licos
- Abiertos sólo para la **aca**demia
- Únicamente para **aut**orizados



Desarrollo de la tecnología lingüística

- 5 situaciones en 2012
  - Situación excelente:
  - Situación adecuada: inglés
  - Situación media: aleman, **castellano**, frances, holandes...
  - Situación en desarrollo: **euskara**, **gallego**, **catalán**, esloveno, sueco, danés…
  - Situación pobre: irlandés (gaélico), islandés, rumano...

En el futuro podremos desarrollar toda la investigación en la nube sin costes adicionales conectándonos desde la infraestructura y explotando los datos lingüísticos en organizaciones diferentes, creando colecciones virtuales, contrastables y desarrollando investigaciones de fácil reproducibilidad.

¿Cómo?

CLARIN-ES en el futuro

Recursos lingüísticos
Language resources

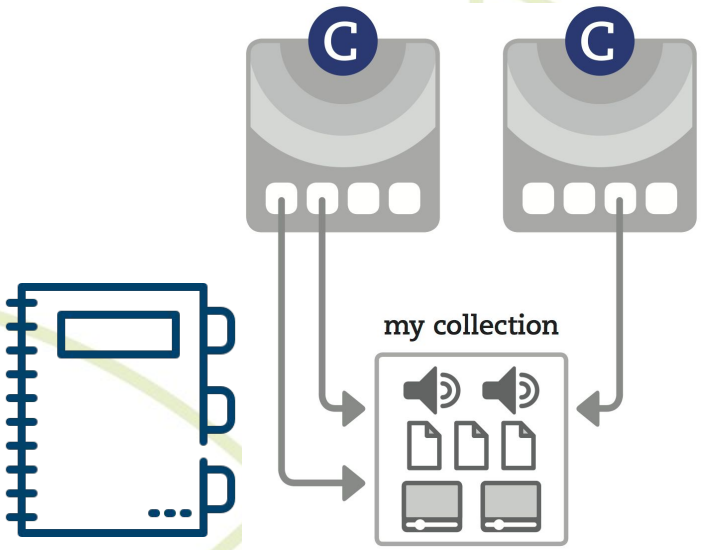
Corpus y metadatos: en grandes cantidades y para búsquedas rápidas:

- vlo.clarin.eu/#tour
- contentsearch.clarin.eu
- https://labur.eus/gZ1ld

Colecciones virtuales
Virtual collections

Para crear corpus virtuales y poder mencionarlos (replicabilidad):

- https://www.clarin.eu/content/virtual-collections



CLARIN-ES en el futuro

Repertorio de herramientas

Language Resource Switchboard

Para encontrar la herramienta adecuada para cada tipo de datos lingüísticos

- https://switchboard.clarin.eu/

3

Depósito recursos

Servicios Depositing services

Para depositar y mantener corpora y recursos:

- www.clarin.eu/content/depositing-services

4

Spanish CLARIN K-centre:
Tecnologías del Lenguaje para eHumanidades y ciencias sociales (y más allá)

www.clarin.eu
www.clarin-es.org

Mikel Iruskieta
Grupo Ixa
HiTZ Basque center for Language Technology
UPV/EHU

CLARIN

eman ta zabal zazu
Universidad del País Vasco
Euskal Herriko Unibertsitatea

HiTZ
Hizkuntza Teknologiako Zentroa
Basque Center for Language Technology