CLARIAH-EUS

Una infraestructura basada en las tecnologías del lenguaje en euskera para las Humanidades y las Ciencias Sociales

Mikel Iruskieta, Jon Alkorta, Begoña Altuna, Xabier Arregi, Jose Mª Arriola, Ainara Estarrona y Aritz Farwell
mikel.iruskieta;jon.alkorta;begona.altuna;xabier.arregi;josemaria.arriola;ainara.estarrona;aritz.farwell@ehu.eus
Grupo Ixa, Centro HiTZ, Universidad del País Vasco/Euskal Herriko Unibertsitatea

¿Qué es CLARIAH-EUS?

CLARIAH-EUS es un proyecto para promover la investigación en euskera en Humanidades y Ciencias Sociales dentro de las infraestructuras de investigación europeas.

¿Por qué CLARIAH-EUS?

- Ofrecer servicios para el euskera al estilo de CLARIN y DARIAH.

Ayuda **Ciencia abierta** **Principios FAIR**
Menos dispersión **Más repercusión**

- Integrar el euskera en las redes de investigación europeas.
- Aumentar los recursos para el euskera para mejorar los resultados de investigación en CC.SS. y Humanidades.
- Unir a todos los investigadores vascos en un espacio compartido.

En euskera en CLARIN

desarrollando las TL para el euskera que no se desarrollaran en Europa.

En euskera en DARIAH-EU

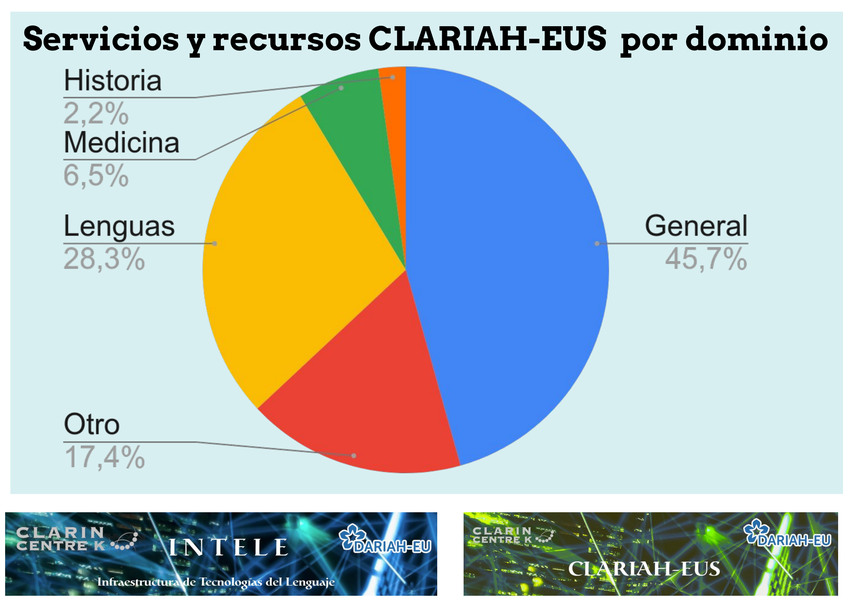
desarrollando recursos, dialogos interdisciplinares en la comunidad cientifica: educación, sociología, historia, GLAM, derecho...

¿Cómo CLARIAH-EUS?

- Plan anual de creación y adaptación de recursos y servicios.
- Evaluación periódica de la infraestructura.

Plan de creación y adaptación de recursos y servicios (5 años)

Año	CLARIN adaptar	DARIAH adaptar	Crear
1	**Servicios**: Incluir servicios de análisis de textos para euskera en la Switchboard: IXAPipes, ANALHITZA... - Incluir servicios de análisis de voz BAS. **Recursos**: EusCrawl y ParlaMint. - Catálogo de tecnologías del lenguaje.	**Servicios**: OCR para el euskara. **Recursos**: Corpus literario, SAHCOBA, redes sociales.	- Sistema para Crawlear, extracción de tuits. - Crear APIs para IXAPipes, ANALHITZA... - Sistema para el alojamiento de proyectos HD. - Recopilación de proyectos HD en la investigación en euskera.
2	**Servicios**: Incluir corpora en el Language Virtual Observatory (VLO). **Recursos**: Modelos de lenguaje.	**Servicios**: Cursos, workshops.	**Proyectos**: Educación, participar en CLARIN Universities. - Creación de corpora y servicios útiles para investigar textos educativos.
3	**Servicios**: workflows para el análisis de textos.	**Servicios/recursos:** adaptar las herramientas al ámbito de sociología. - Cursos, workshops.	**Proyectos**: Sociología y periodismo. - Creación de corpus y servicios útiles para investigar textos educativos.
4	**Servicios**: casos de uso, traductor automático.	**Servicios/recursos para GLAM:** Cursos, workshops.	**Proyecto**: Literatura y GLAM.
5	**Servicios/recursos:** casos de uso, catálogo de recursos y servicios.	**Servicios/recursos para el amb. de Historia**: Cursos, workshops.	**Proyecto:** Historia. - Recopilación de proyectos HD en la investigación en euskera.



Evaluación de CLARIAH-EUS

- Estadísticas de uso de la infraestructura.
- Evaluación del nivel de satisfacción de las personas usuarias.
- Medición del impacto en investigación y en ámbitos social, público y económico.

Casos de uso

Parlamint-EUS y BasqueParl

Textos parlamentarios

¿Cómo se usa el code switching euskera-castellano en el Parlamento Vasco?

VaxxStance

Redes sociales

¿Están los tuiteros a favor o en contra de las vacunas?

- Microblogs en ES, EU.
- Selección por palabras clave: vacuna, #pfizer…
- Anotación manual.

European Literary Text Collection, ELTeC

Corpus literario

- 26 novelas, s. XIX-XX, euskera.
- Distant reading.
- Corpus comparables en lenguas europeas.

BIM/SAHCOBA

Textos históricos

- Objetivo: sintaxis diacrónica.
- Textos s. XV-XVIII/XVIII-XX.
- Análisis morfosintáctico.
- Interfaz de búsqueda de estructuras.

Aprendizaje del euskera

Educación

- Acuerdos con HABE, Elkar y Glifing.
- 0-22 años, niveles A1-C2.
- Clasificación por competencias lingüísticas.