CLARIAH-EUS Humanitateetako eta Gizarte Zientzietako euskarazko ikerketa-azpiegitura marrazten HaMABI printzioei jarraituz

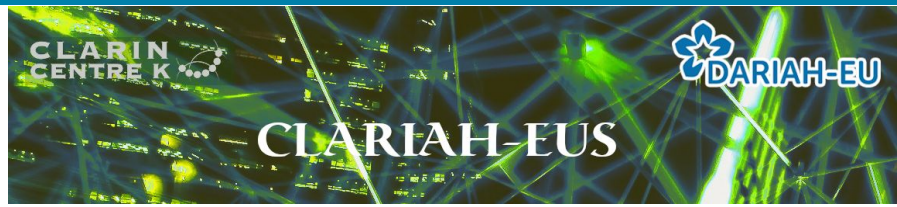
www.clarin.eu
www.dariah.eu

http://ixa2.si.ehu.eus/intele/

Ainara Estarrona
INTELE - HiTZ (UPV/EHU)
Mikel Iruskieta
INTELE - HiTZ (UPV/EHU)

eman ta zabal zazu
Universidad del País Vasco
Euskal Herriko Unibertsitatea
HiTZ
Hizkuntza Teknologiako Zentroa
Basque Center for Language Technology

Aurkezpenaren egitura



1. Sarrera
2. INTELE ikerketa-sarearen aurkezpena
3. Europako ikerketa-azpiegituren aurkezpena: CLARIN eta DARIAH
4. CLARIN-K zentroaren aurkezpena: erabiltzaileen esperientziak
5. Ikerketa errazten: kasu praktikoak
6. Eztabaida eta galderak

CLARIN

DARIAH-EU

SAIOAREN LABURPENA: CLARIAH-EUS ZERTARAKO ETA NOLA?

Helburua

- *European Open Science Cloud*: EOSC
- HaMABI printzipioen zientzia (FAIR): Hartu, Moldatu, Aztertu, Bistaratu eta Iraunarazi

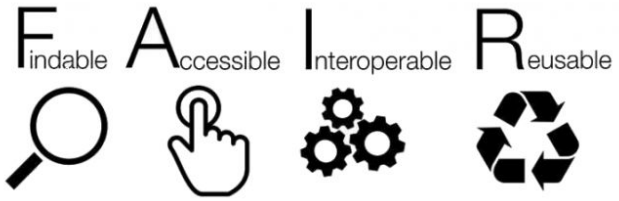
Metodoa

- Elkarreragingarritasuna azpiegituretan (CLARIN)
  - Europan garatuko ez den euskararako azpiegitura eraiki
  - ALL-LT-in-ONE-URL: baliabide guztiak, zerbitzu guztiak

Adibideak

- Elkarren artean komunikatzen diren azpiegitura europarretan ikertzeko erabilera kasuak eta tresnak (erabilerrazak)

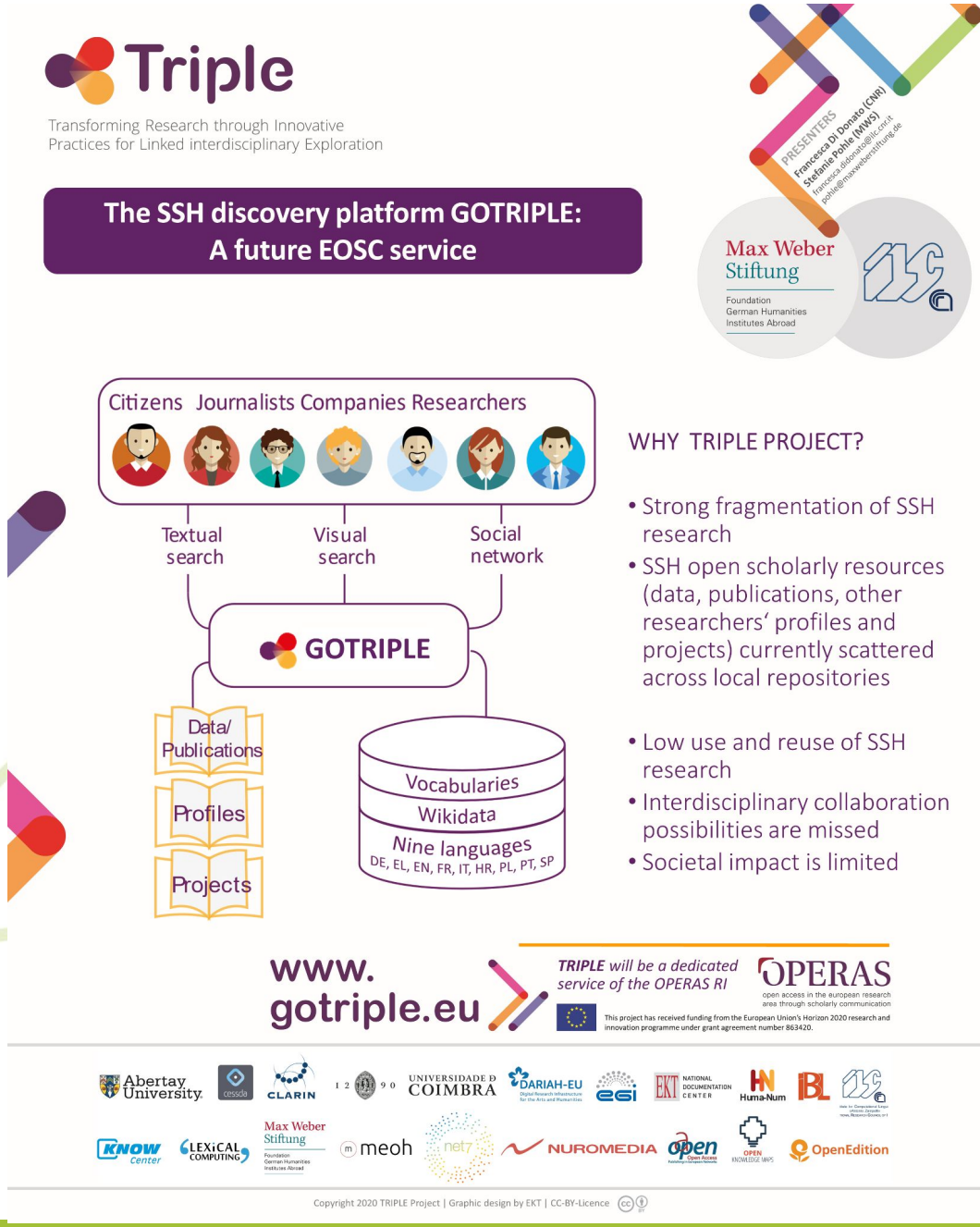
EUROPEAN OPEN SCIENCE CLOUD



Azpiegituren justifikazioa

Zientzia: irekia, eraginkorra eta errepikagarria:

- Humanitateetako eta Gizarte Zientzietako ikerketaren zatiketa
- Humanitateetako eta Gizarte Zientzietako baliabideak biltegi desberdintean sakabanatuta
- Gizarte Zientzietan ikerketaren berrerabilpen gutxi
- Diziplinartekotasun gutxi
- Eragin sozial mugatua



INTELE: Red estratégica para la promoción de las **in**fraestructuras de **te**cnologías del **le**nguaje en eHumanidades y Ciencias Sociales

- **INTELE**: CLARIN eta DARIAH europar azpiegituretan Espainiaren parte hartze ofiziala bultzatzeko ikerketa-sare estrategikoa
  - Humanitateetan eta Gizarte Zientzietan ikerketa sustatzea
  - Nazioarteko proiektuak eta programak sustatzea
  - Helburua: CLARIAH-ES
- **CLARIN: C**ommon **LA**nguage **R**esources and Technology **IN**frastructure
  - ESFRI ERIC (2012) y ESFRI Landmark (2016)
- **DARIAH: D**igital **R**esearch **I**nfrastructure for the **A**rts and **H**umanities

Strategy Report on Research Infrastructures ROADMAP 2018 CLARIN DARIAH-EU

INTELE ikerketa-sarea

- Parte hartzaileak
  - 9 unibertsitate eta ikerketa zentro
    - **EHU** (HiTZ), UPF, UVigo, UNED, UCM, UAlicante (BVC), UJAEN, USC (Instituto da Lingua Galega y CiTIUS) eta CSIC (Centro de Ciencias Humanas y Sociales, etc.)
- Babesa:
    - BNE, RAE, etc.
    - 35 unibertsitate baino gehiago eta 150 ikertalde eta ikerketa-sare baino gehiago
- Ekimenak:
    - 2020: Aurkezpen workshopa (online): 130 lagun baino gehiago
    - 2021: Nazioarteko web mintegiak (hilero)
    - 2021: Alor desberdinetako adituen workshopa
      - Hizkuntzalaritza, historia, zuzenbidea, liburutegiak, hezkuntza, etab.

https://ixa2.si.ehu.eus/intele/?q=home

INTELE G1: Historia +

- **Teknika metodologiko kualitatiboetarako tresnak**:
  - Edukien analisia, testuena, diskurtsoarena, irudiena, kazetaritza testuen analisia, testu meatzaritza, sentimenduen analisia (sare sozialak), web scraping-a, ahozko iturriak, elkarrizketak, elkarrizketen transkripzioa, testuen sailkapena, dokumentuen etiketatzea, eztabaida-taldeak, etnografia, azterketa biografikoak
- **Teknika metodologiko kuantitatiboetarako tresnak**:
  - Inkestak, datu-baseak, estatistikak, testu corpusak, datuen meatzaritza (ikusentzunezko baliabideak eta baliabide digitalak), audientzien azterketa, big data, sare sozialak
- **Bistaratze eta lokalizazio tresnak**:
  - Informazio geografikoa: infografiak, mapa irudiztatuak, grafikak, mapa bizidunak, zuhaitz genealogikoak, arazo eta konponbideen zuhaitzak, sare sozialetatik eratorritako grafoak
- **Edukien biltegiratzea:**
  - Iturriak: ikusentzunezkoak eta digitalak, datuen babesa eta ikerketaren kontserbazioa
- **Zerbitzuak**:
  - Formazioa, elkarreragingarritasuna, datu irekiak, itzultzaile automatikoak, konponbide informatikoen katalogoa
- **Datuak**:
  - Eduki irekiak, datuak deskargatzea modu malgu eta errazean

Europako ikerketa-ekosistema digitala

EUROPEAN OPEN SCIENCE CLOUD

SSHOC
social sciences & humanities open cloud

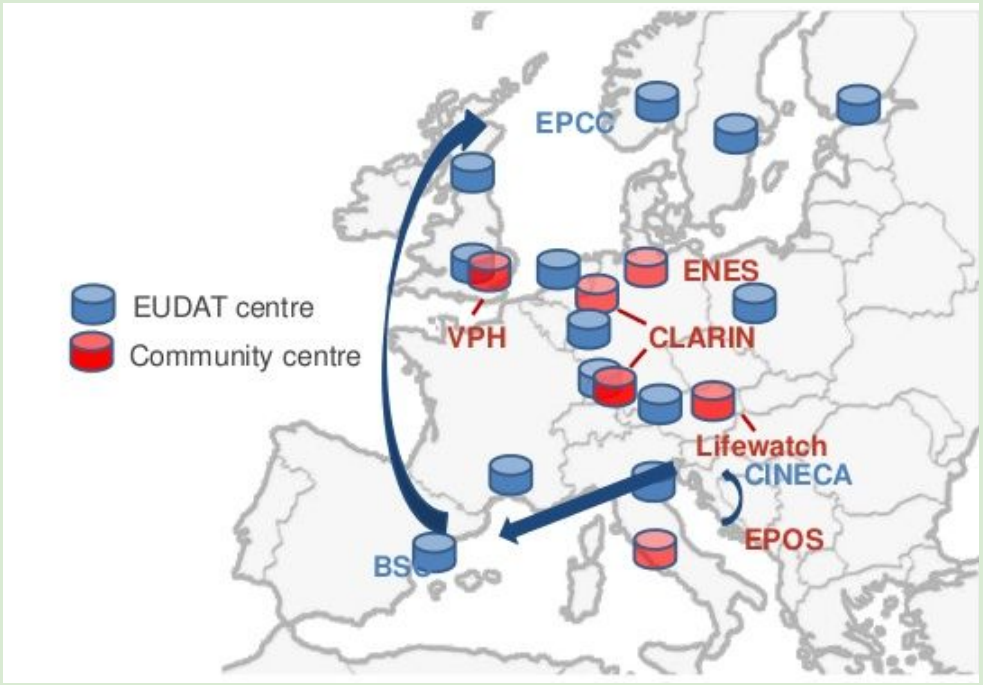
FEDERATED DATA INFRASTRUCTURE

PRACE

OpenAIRE

EUROPEAN LANGUAGE GRID

EUDAT



https://www.slideshare.net/EUDAT/b2-safe-how-to-replicate-your-data

Elkarreragingarritasuna!

SSHOC: Humanitate Digitaletako eta Gizarte Zientzietako e-azpiegituren konexioa

ESFRI - SSHOC

CESSDA ERIC - Consortium of European Social Science Data Archives

CLARIN ERIC - Common Language Resources and Technology Infrastructure

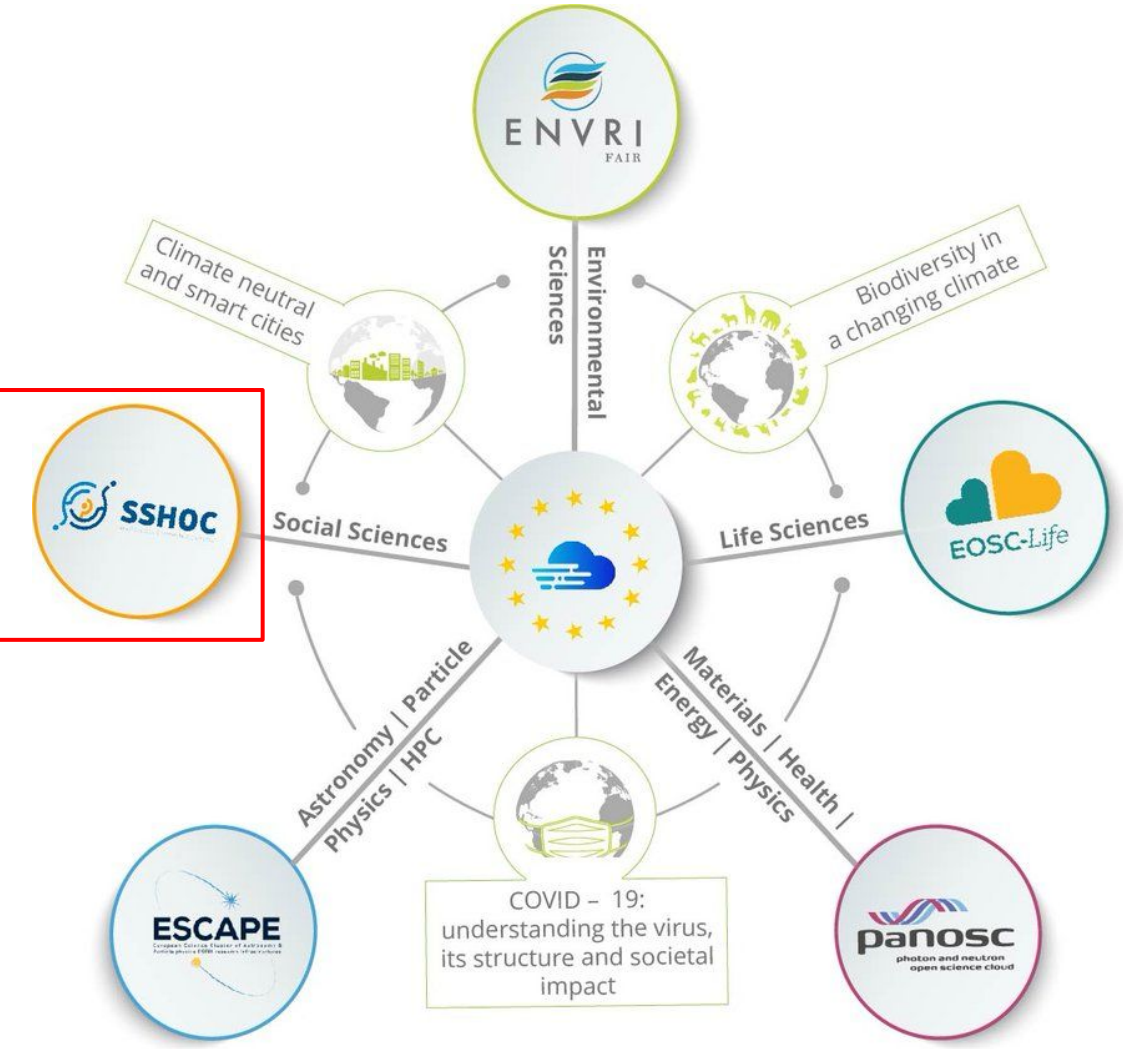
DARIAH ERIC - Digital Research Infrastructure for the Arts and Humanities

ESS ERIC - European Social Survey

SHARE ERIC - Survey of Health, Aging and Retirement in Europe

E-RIHS - European RI for Heritage Science - CSIC, CENIEH

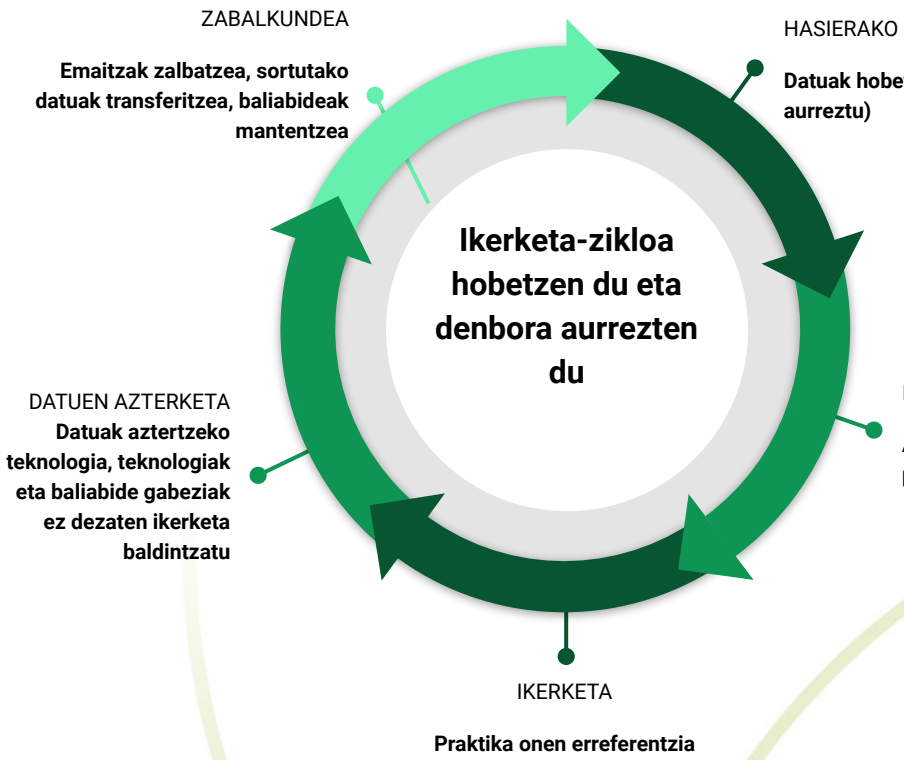
EOSC azpiegituren mapa



ESFRI European Strategy Forum on Research Infrastructures

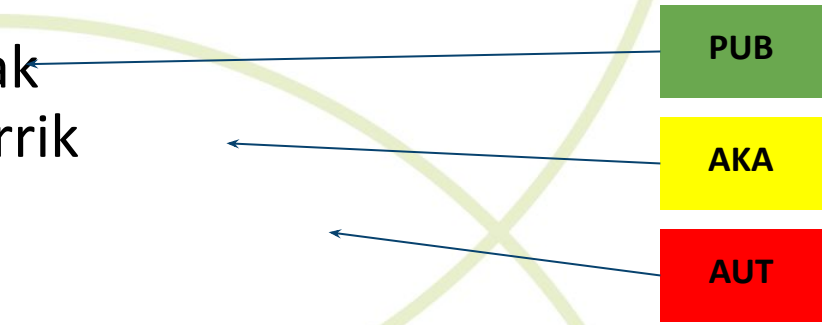
#SSHOC Europako e-azpiegiturak ***European Open Science Cloud*** delakoan sendotzeko eta konektatzeko proiektua da.

Azpiegitura bat proiektu bat baino askoz gehiago da



- Sarbide konfederatua baliabide eta datu guztietara web-gune bakar batetik
- **Estandarrak**
- **Protokolo komunak**
- Paradigma aldaketarako laguntza
- Baliabide **estrategikoen** diseinua

- Corpusak: irekiak eta publikoak
- Irekiak akademiarentzat bakarrik
- Baimenduentzako bakarrik



Language Resource Switchboard

Zure hizkuntza datuentzat egokien den tresna aurkitzeko:

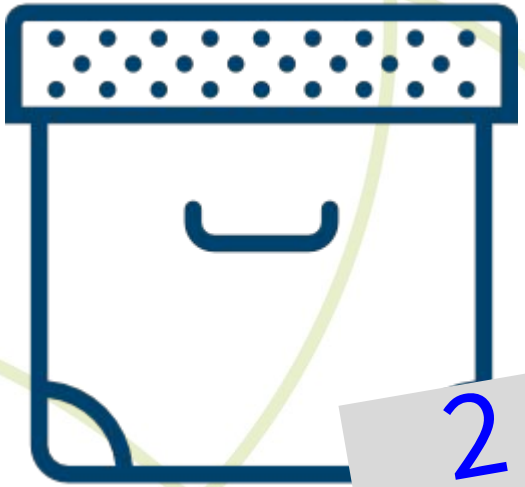
- https://switchboard.clarin.eu/

1

Depositing services

Corpusak eta baliabideak uzteko eta mantentzeko:

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

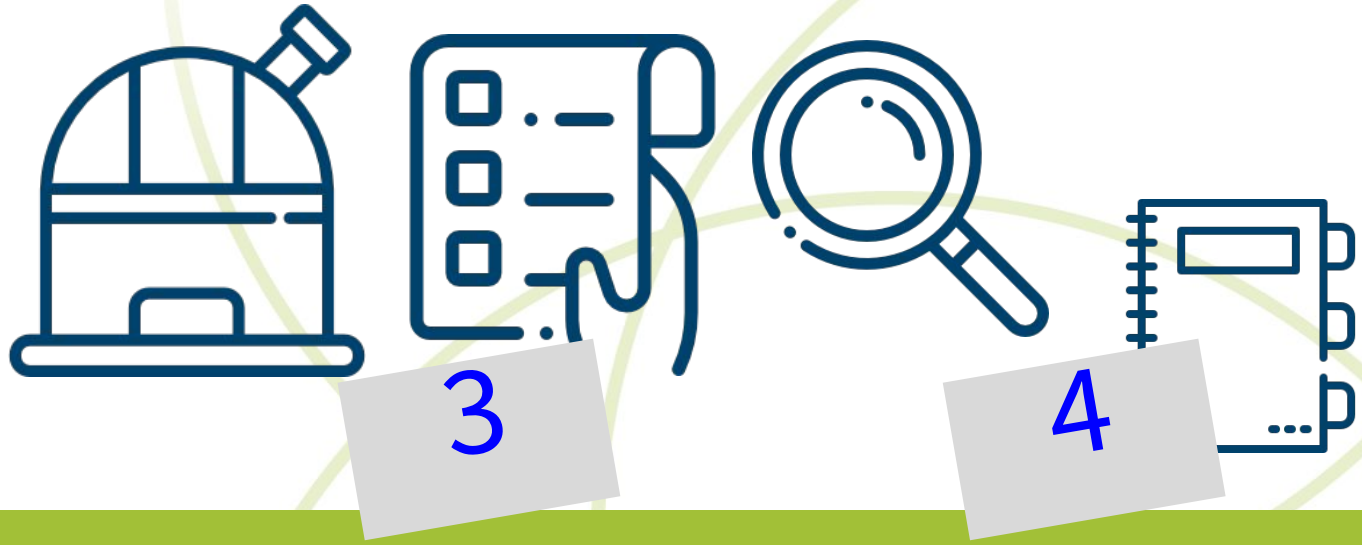


2

Language resources

Corpusak eta metadatuak: kopuru handiak eta bilaketa azkarrak:

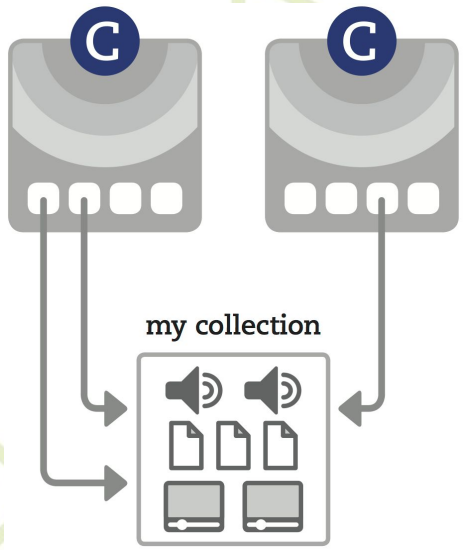
- vlo.clarin.eu/#tour
- contentsearch.clarin.eu
- https://labur.eus/gZ1ld



Virtual collections

Corpus birtualak sortu eta horiek aipatu ahal izateko (erreplikagarritasuna):

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

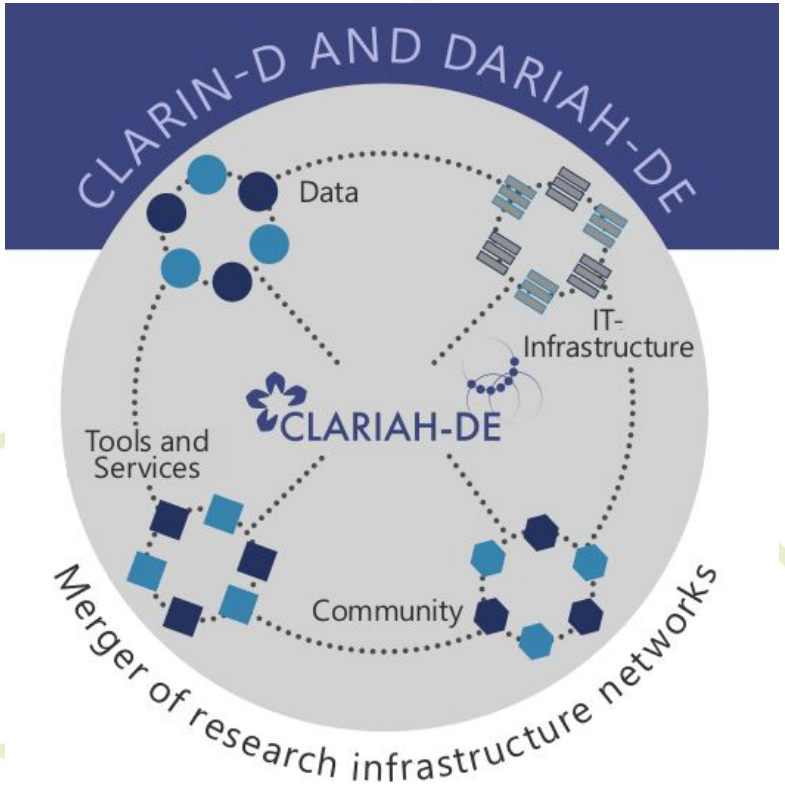


Azpiegitura eredua: CLARIAH (CLARIN + DARIAH)

ESFRI European Strategy Forum on Research Infrastructures

THE NETWORK IN EUROPE

The CLARIN, DARIAH and CLARIAH research infrastructures are active on a national and the European level. CLARIN and DARIAH both have the status of a European Research Infrastructure Consortium (ERIC).



The content-related and technological foundations created by CLARIN-D and DARIAH-DE will be aligned, integrated, further developed and jointly maintained in CLARIAH-DE.

Hizkuntzen egoera CLARIN teknologian

https://vlo.clarin.eu milioi bat elementu baino gehiago daude (tresnak, corpusak...)

- Ingelesa (154.928)
- Alemana (143.271)
- Nederlandera (117.312)
- Daniera (109.962)
- Esloveniera (73.495)
- Poloniera (40.466)
- Frantsesa (24.432)
- Afrikaansa (7.870)
- ...

CLARIN-es hizkuntzetako baliabideak

- Gaztelania (14.444)
- Katalana (1.364)
- **Euskara** (498)
- Galegoa (216)

- Helburua eta konpromisoa
  - HaMABI
  - Garapen jasangarria

=

ALL LTs in ONE url

DATA TYPE	
spoken	99558
speech	4455
writing	1543
gestures	1307
pointing-gestures	454
facial-expressions	452
emotional-state	451

Nederlanderaren adibidea

CLARIAH
Common Lab Research Infrastructure for the Arts and Humanities

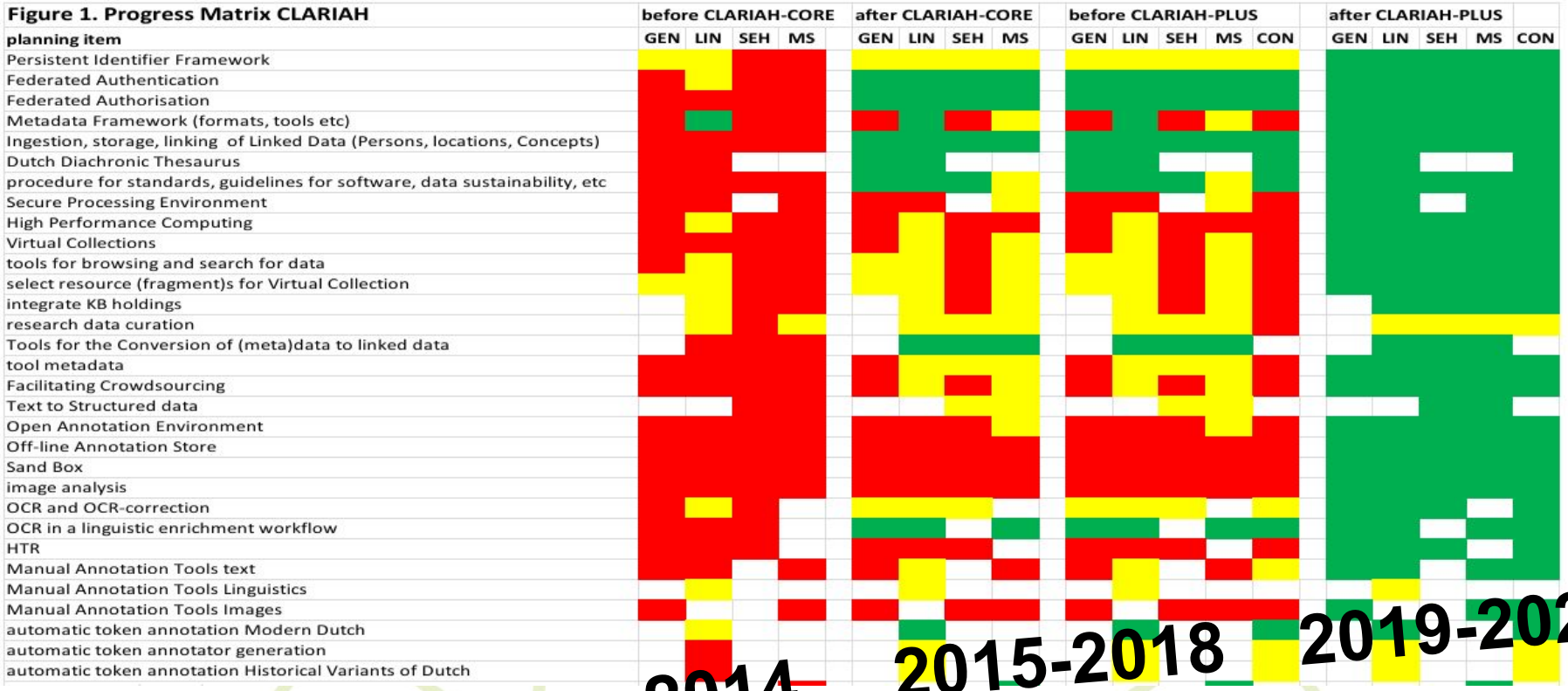
National Roadmap for Large-Scale Research Infrastructure

1 General information

GENeric functionality, LINguistics, Socio-Economic History and Media Studies, CONtent of texts: history, literary

Iturria:

https://www.clariah.nl/over/bestanden/downloads/send/10-folders/166-clariah-plus



CLARIAH-EUS proposamena

- CLARIN eta DARIAH batzea
- Antolaketa eta finantziazioa
  - Unibertsitateak, ikerketa zentroak, erakundeak, etab.
  - Eusko Jaurlaritza, aldundiak, udalak, etab.
- Etorkizuneko CLARIAH-EUS azpiegituraren ekimenak
  - a) **Ikertzaileen koordinazioa eta zerbitzuak**
    - > 2 lagun (FTE)
  - b) **Zerbitzuak + Zabalkundea + Formazioa**
    - > Lagun 1 (FTE)
  - c) **Teknologia berriaren garapena**
    - > Lagun 1 (FTE), betiere garatu beharreko proiektu eta ekintzen arabera

CLARIAH-EUS: Bide-orria

- Proiektua aurkeztu
  - Unibertsitateetan, ikerketaldeetan, ikerketa-zentroetan, euskal erakundeetan, instituzioetan, etab.

- CLARIAH-EUS partzuergoaren diseinua osatu

- **CLARIAH-EUS** CLARIN eta DARIAH azpiegituretan nodo proaktiboa izan

CLARIN-K zentroa eta bere nodoak

http://clarin-es.org

1. IXA-EHU: http://ixa2.si.ehu.eus/clarink/index.php?lang=es
2. IULA-UPF: http://www.clarin-cat-lab.org/index-en.html & http://services.iula.upf.edu/
3. LINDH-UNED: https://linhd.uned.es/clarin-centre-k/
4. TALG-UVIGO: http://sli.uvigo.gal/clarink/

CLARIN: erabiltzaileen esperientziak

https://zenodo.org/record/4288980#.X9YDS7N7mDI

CLARIN-K zentroaren ad hoc zerbitzua

CLARINeko Tour delakoan argitaratua

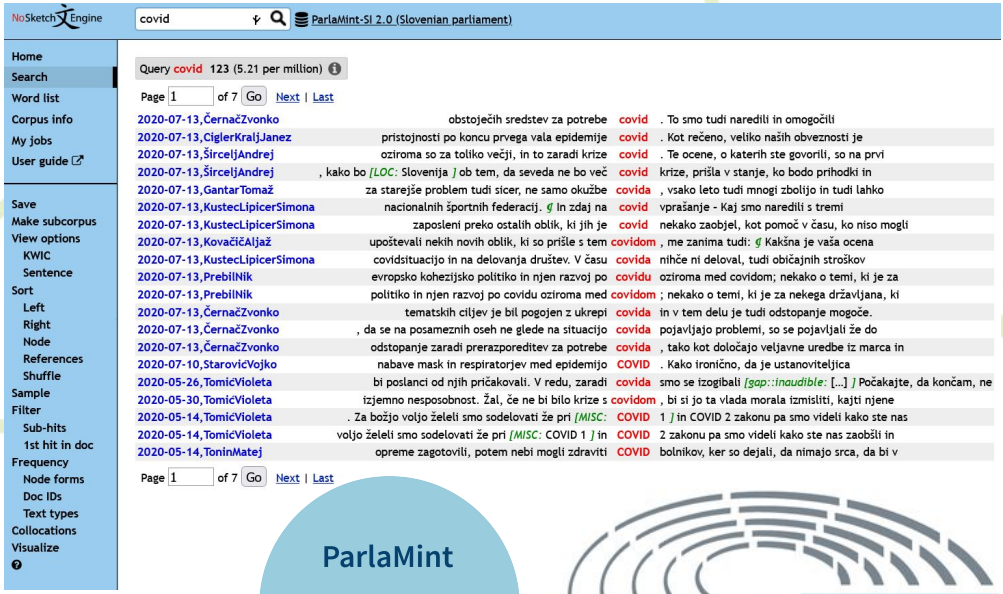
- Tesia *Basque Center on Cognition, Brain and Language* (BCBL) ikerketa-zentroan egiten du
- Gaia: “My PhD work focuses on the amount of exposure to each language within bilingual contexts, and how it shapes language acquisition at a cognitive and neural level”
  - **Tresnak**:
  - ANALHITZA
  - https://switchboard.clarin.eu

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

ParlaMint corpus eleanitza(testu idatzia)

1. Legebiltzarretako datuak eta metadatuak lortzea
2. ParlaMint eskemara bihurtzea
3. Anotazio linguistikoa (UDpipe eta NERC)
4. Bat datozen corpusen bidez corpus berri erabilgarriak sortzea (noSketch Engine / KonText) eta Parlameter

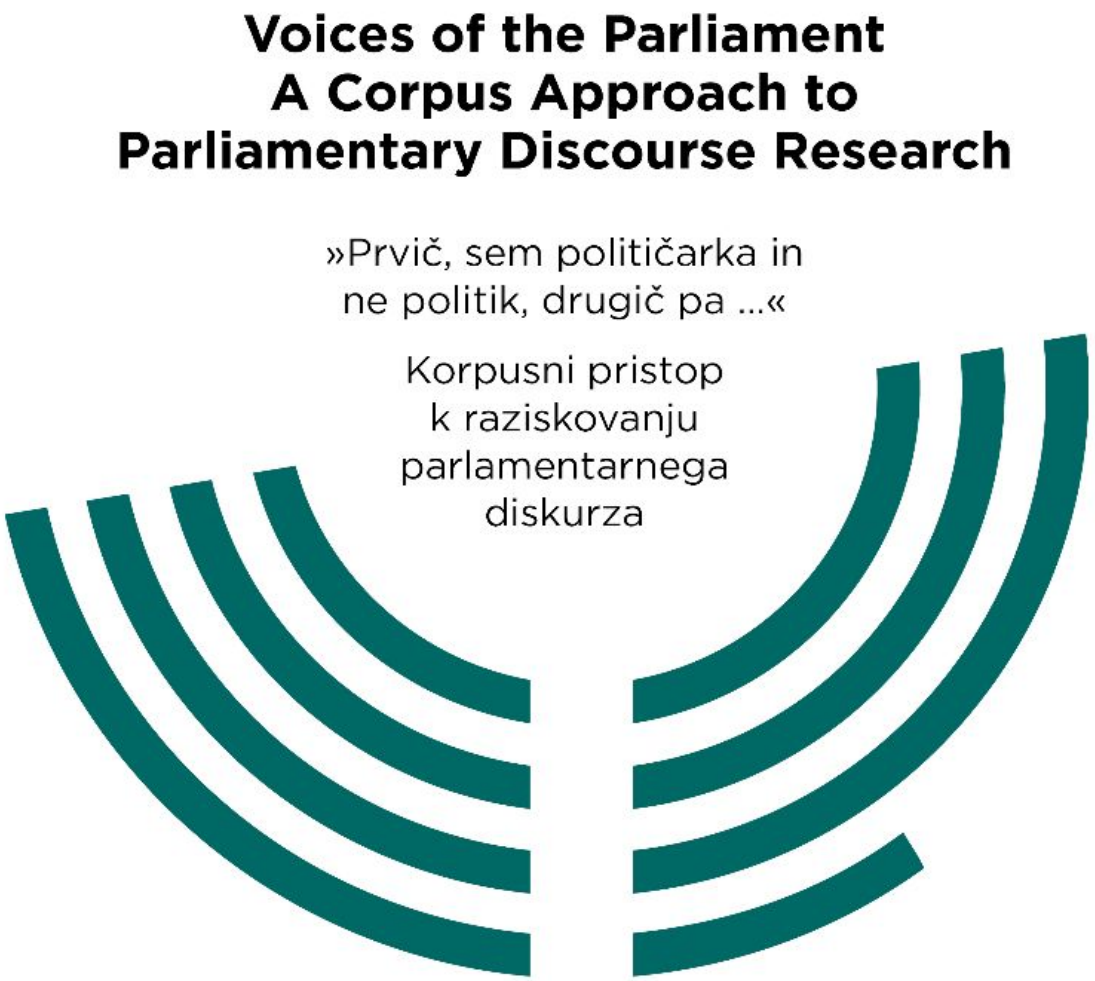
https://www.clarin.si/noske/parlamint.cgi/



INTELEko web mintegian ikerketa-galderak: https://youtu.be/b0oNElZbV9E

Corpusa modu praktikoan erabiltzeko materiala

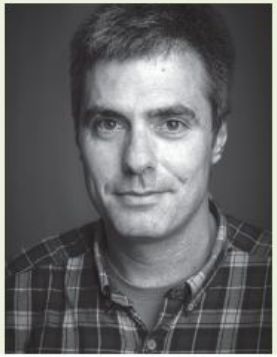
https://sidih.github.io/voices/toc.html



CLARIN-K IMPACT-CKC zentroaren erabilera

IMPACT-CKC K-CENTRE

Interview | **Mikel Iruskieta**



Mikel Iruskieta is a computational linguist who is part of the Ixa Research Group and the Didactics of Language and Literature Department at the University of the Basque country. He has collaborated with the CLARIN IMPACT-CKC Knowledge Centre, which helped him and his colleagues digitize Basque texts.

Could you briefly describe your academic and research background?

<

My current research focuses on the didactics and analysis of Basque, mostly regarding discourse parsing and evaluation of discourse structure. For the last five years, I have mainly worked on adapting language technologies for teaching and learning purposes. With that goal, I have created and now co-lead a postgraduate programme in Basque (University Specialist in ICT and Digital Competences in Education, Continuing Education and Language Teaching), a research group working in Digital Humanities and Education. Our aim is to build a research community that will conduct research and teach in Basque by adopting a critical approach and using language technologies in a pedagogical context. In this postgraduate programme, my colleagues and I are developing a new framework of the socio-tech pedagogy for Basque that will cover the following topics:

- The Basics of Technology and Pedagogy;
- Formal Education and Technology;
- Continuing Education and Technology;
- Language Teaching and Technology Development;
- Society and Education, Opportunities and Risk of Technology;
- E-learning: Approaches and resources; and
- Digital Research: Methods and resources.

>

Does the fact that Basque is a language isolate have any bearing on the development of language tools tailored to it?

<

The history and current situation of the Basque language are both complex and interesting. Basque has a relatively small community of speakers (751,700 active and 1,185,500 passive speakers) which lives in contact with three powerful language communities, namely Spanish and French (as official languages in the Basque Country) and English (as a foreign language). It is also not supported enough by official language policies. As a result, Basque is still considered an under-resourced language. In this context, the work of the Ixa Group for NLP is highly valuable. They have developed basic resources for Basque (as well as for other languages) which are used by the research community, for example IXApipes (a modular set of NLP tools which provide easy access to NLP technology for several languages that can be used or exploit its modularity to pick and change different components) and ANALHITZA (a web service to analyse Basque, Spanish and English texts without needing any technical experience). Many more basic and advanced tools and resources for Basque can be found on the website of the HiTZ: Basque Center for Language Technology.

>

How did you get involved with the IMPACT K-Centre and how did they help you with your research?

<

I learned about the IMPACT K-Centre when they joined CLARIN. Because I was working on several different digitization projects for Basque and for Spanish, I immediately got in touch with them and asked for their help. Isabel Martínez Sempere, the manager of IMPACT, helped me solve a digitization issue that I encountered when I was analysing the most frequently occurring words in *Pulgarcito*, which is a Cuban children's magazine written in Spanish from 1919 to 1920. This magazine consists of very diverse materials, such as drawings and handwritten texts, which are

http://clarin-es.org/tour-de-clarin-vol-iii/

IMPACT CLARIN K-centre eta BNE

impact digitisation.eu centre of competence

Vida de Lazarillo de Tormes

IMPACT DATASET BROWSER

This resource is property of BIBLIOTECA NACIONAL DE ESPAÑA BNE

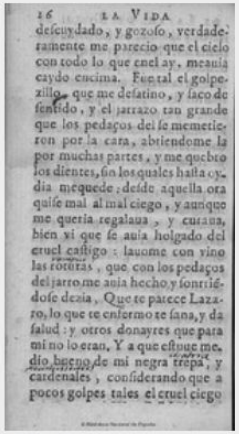
and distributed by the Impact Centre of Competence

impact digitisation.eu centre of competence

ABOUT TOOLS & RESOURCES KNOWLEDGE COMMUNITY CKC HELPDESK

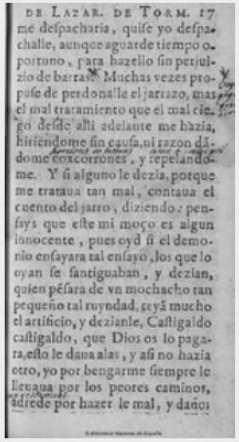
440435

TIFF XML



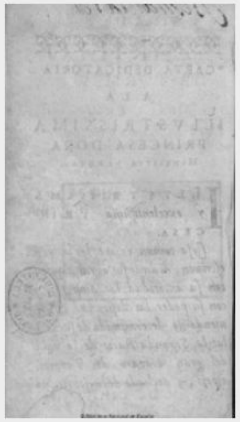
440436

TIFF XML



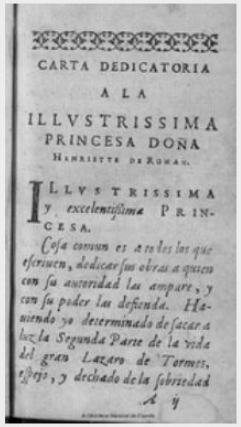
440437

TIFF XML



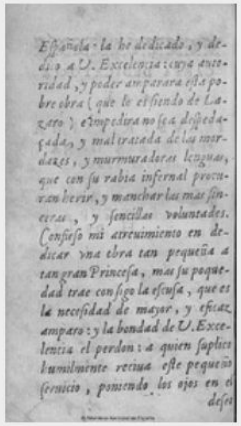
440438

TIFF XML



440439

TIFF XML



440440

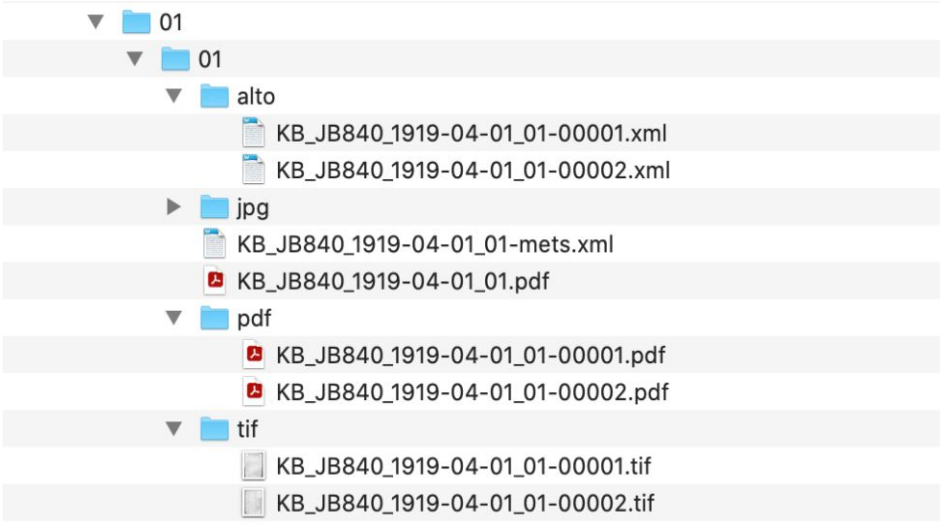
TIFF XML

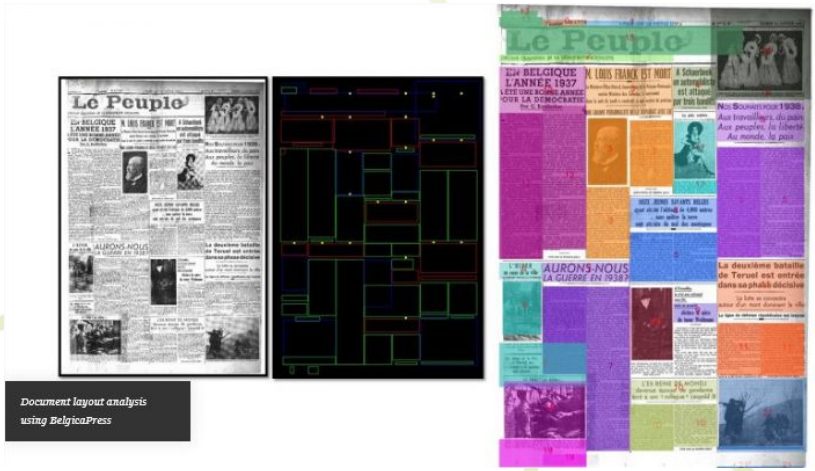
440441

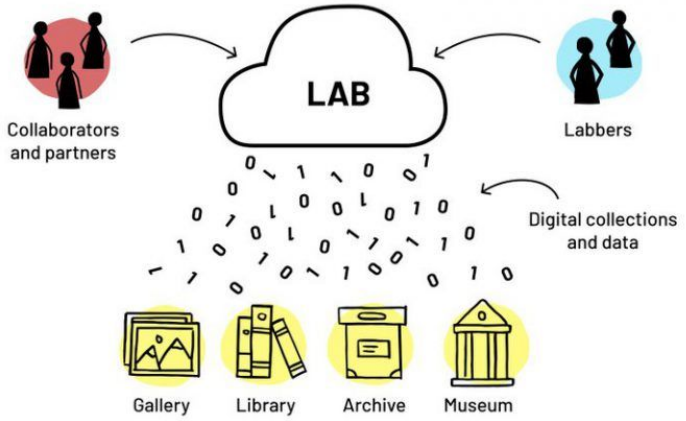
TIFF XML

DATA-KBR-BE: data as collection. DARIAH

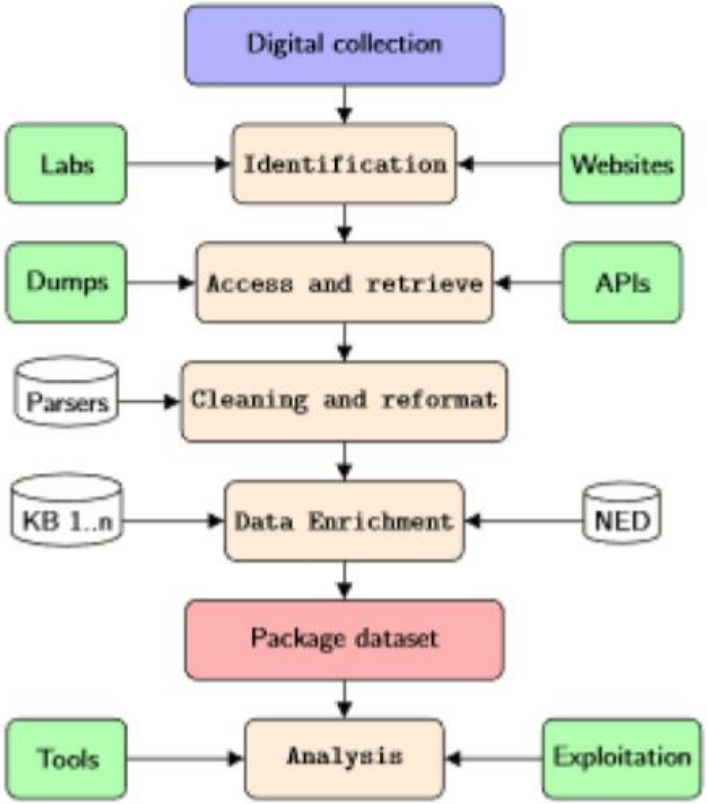
- **Datuak eskura jarri eta HDetan ikertzeko edizio digitalak sortu**
  - Datu egokiak lortzeko lan-fluxua diseinatu
  - Open Data plataforma diseinatu
  - Bilduma digitalen inbentarioa
  - Datasetak argitaratu
  - Hackathon bat datasetak erabiliz







Datuak bildumatzat. BVMC



INTELEko web mintegia: "Bilduma digitaletarako sarbide konputazionala erraztuz". (Biblioteca Virtual Miguel de Cervantes)

Saio praktikoa: github.com/hibernator11/notebook-ph

SSHOC: Train-the-Trainer Bootcamp for Librarians: https://zenodo.org/record/3970799#.YMn2gqZ7mL1

Erabilera-kasuak:
HaMABI printzipioekin lainoan ikerketa errazten CLARIN eta DARIAH azpiegituretan

Azpiegiturak erabiltzen

Digitalizaziorako tresnak

1. Eudat: datuak gorde eta partekatu >>> analizatu

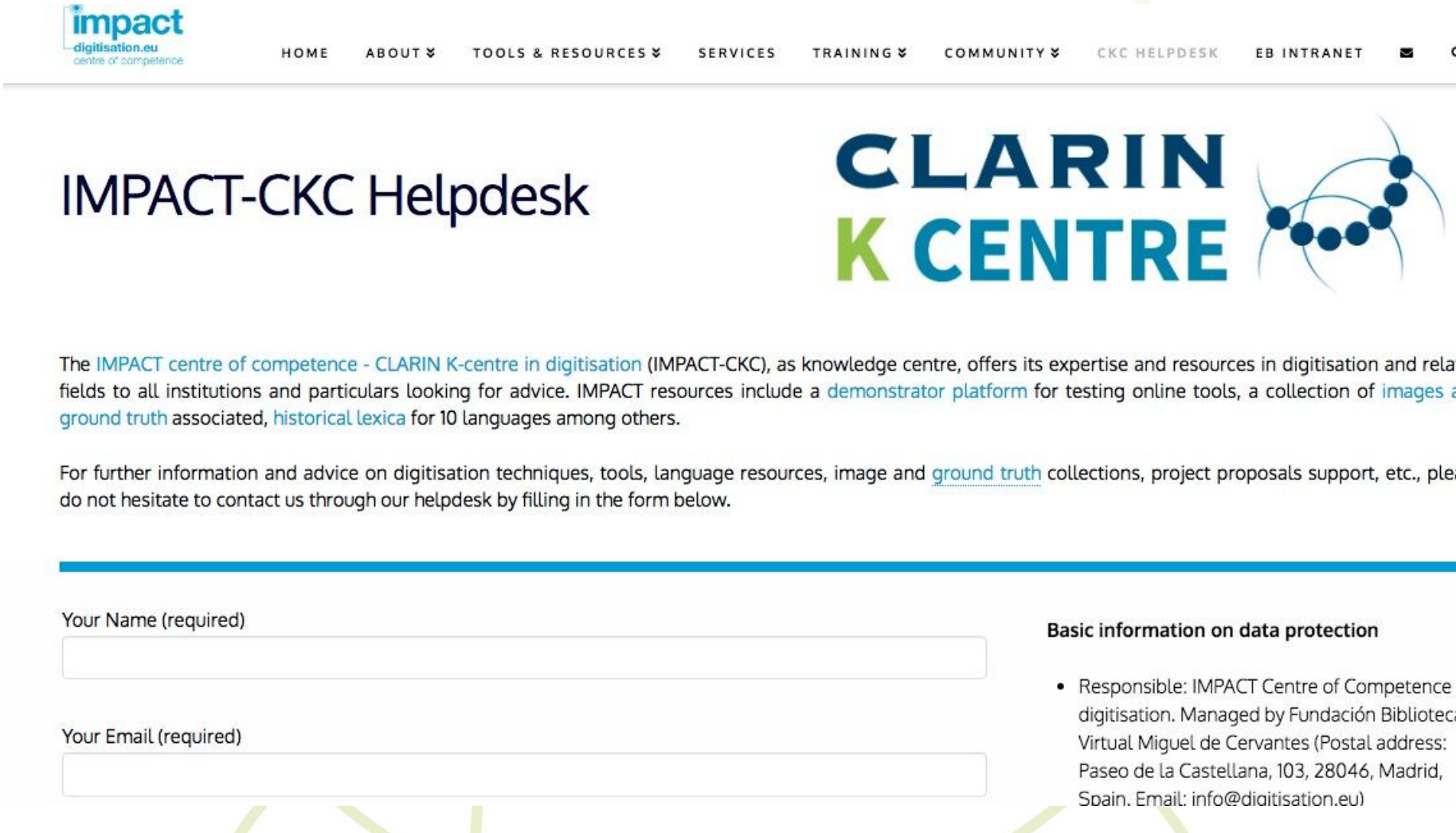
Testu idatzia:

2. Eudat: FAIR data across borders and disciplines
3. Elkarreragingarria CLARINeko Switchboardarekin
4. Hizkuntza detekzioa eta tresna/metodoen aukera
5. Testua erraz aztertu eta bistaraatu/interaktuatu

Ahotsetik testura:

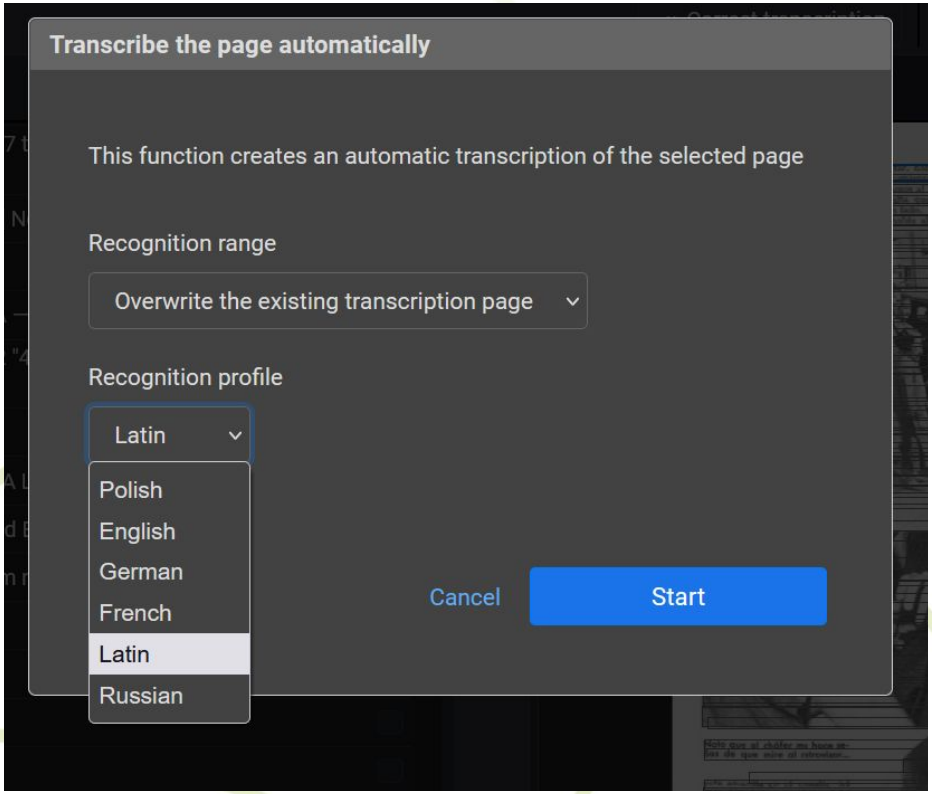
6. BAS zerbitzua CLARINen
7. Hamaika hizkuntza eta aldaera
8. Hamaika irteera formatu, ikertzen jarraitzeko: TXT, SCV, PRAAT, Video...

IMPACT beste CLARIN-K zentrua eta CLARINeko nodoen arteko lankidetza



https://www.clarin.eu/blog/tour-de-clarin-impact-ckc-k-centre

Laborategi birtuala: https://wlt.pcss.pl/



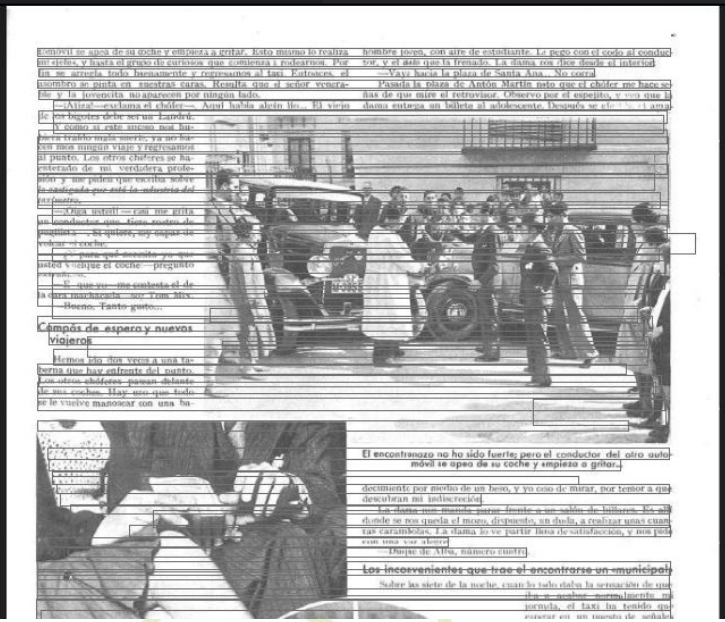
Tresna automatikoak eta eskuzko orrazketa

No.	Page name	Number of rows	Verified	Correct	Incorrect	Last edition	Verification
1	Seis_horas_dentro_de_un_taxi_Carranque_p_1.pdf	87	0	0	0	2021-11-15 10:29	Verify
2	Seis_horas_dentro_de_un_taxi_Carranque_p.2.pdf	87	0	0	0	mikel.iruskieta@ehu.eus 2021-11-15 10:56	Verify

Fold

Transcribed object

Edit selection



Fold

Transcription

1	.
2	tomóvil se apea de su coche y empieza a gritar. Esto mismolo realiza | hombre joven, con aire de estudiante. Le pego con el codo al conduc-
3	mi «jefe», y hasta el grupo de curiosos que comienza a rodearnos. Por tor, y el auto que.la frenado. La dama. nos dice desde el interior:
4	fin se arregla todo buenamente y regresamos al taxi. Entoaces, el — "Vaya hacia la plaza de Santa Ana... No corra.
5	asombro so pinta en. nuestras caras. Resulta que cel seüor vencra- Pasada la plaza de Antón Martín noto que el chófer mo hace sc-
6	ble y la jovencita no aparecen por ningün lado. nas de que mire el retrovisor. Observo por el espejito, y «o que la
7	—¡Atiza!—exclama el chófer—. Aquí había algún lío.. El viejo — dama entrega un billete al adolescente, Después se efe t; i1 agia-
8	de los bigotes debe ser un Landrü. BOB - yep ni xi cERdiacct eras SESS *

Monitorizaziorako tresnak

Title: Seis horas dentro de un taxi

Author: Andrés Carranque de Ríos

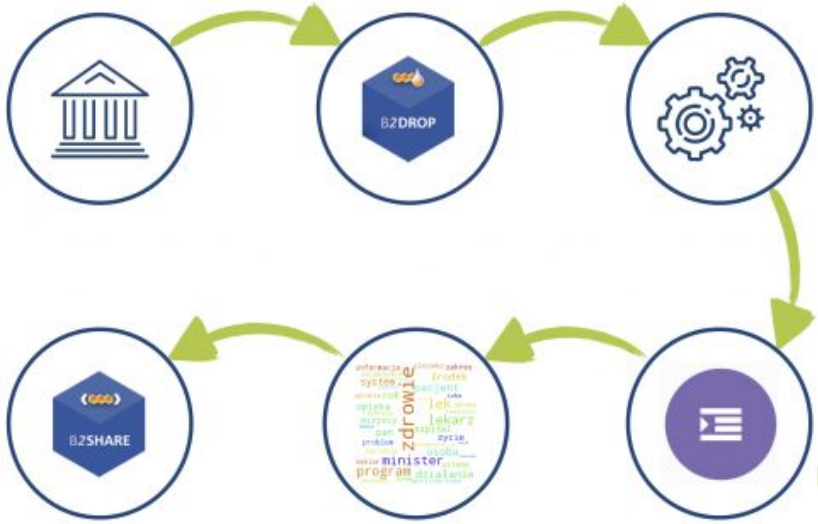
Advanced verification: Pages to verify at this property: 1 Verified rows: 87 / 174 Correct: 83 Incorrect: 4

No.	Page name	Number of rows	Verified	Correct	Incorrect	Last edition	Verification
1	 Seis_horas_dentro_de_un_taxi_Carranque_p_1.pdf	87	0	0	0	2021-11-15 10:29	Verify
2	 Seis_horas_dentro_de_un_taxi_Carranque_p.2.pdf	87	87	83	4	mikel.iruskieta@ehu.eus 2021-11-15 10:59	Verify

Erabilera-kasua CLARINen: HaMABI printzipioak

EuDAT azpiegitura erabiliz

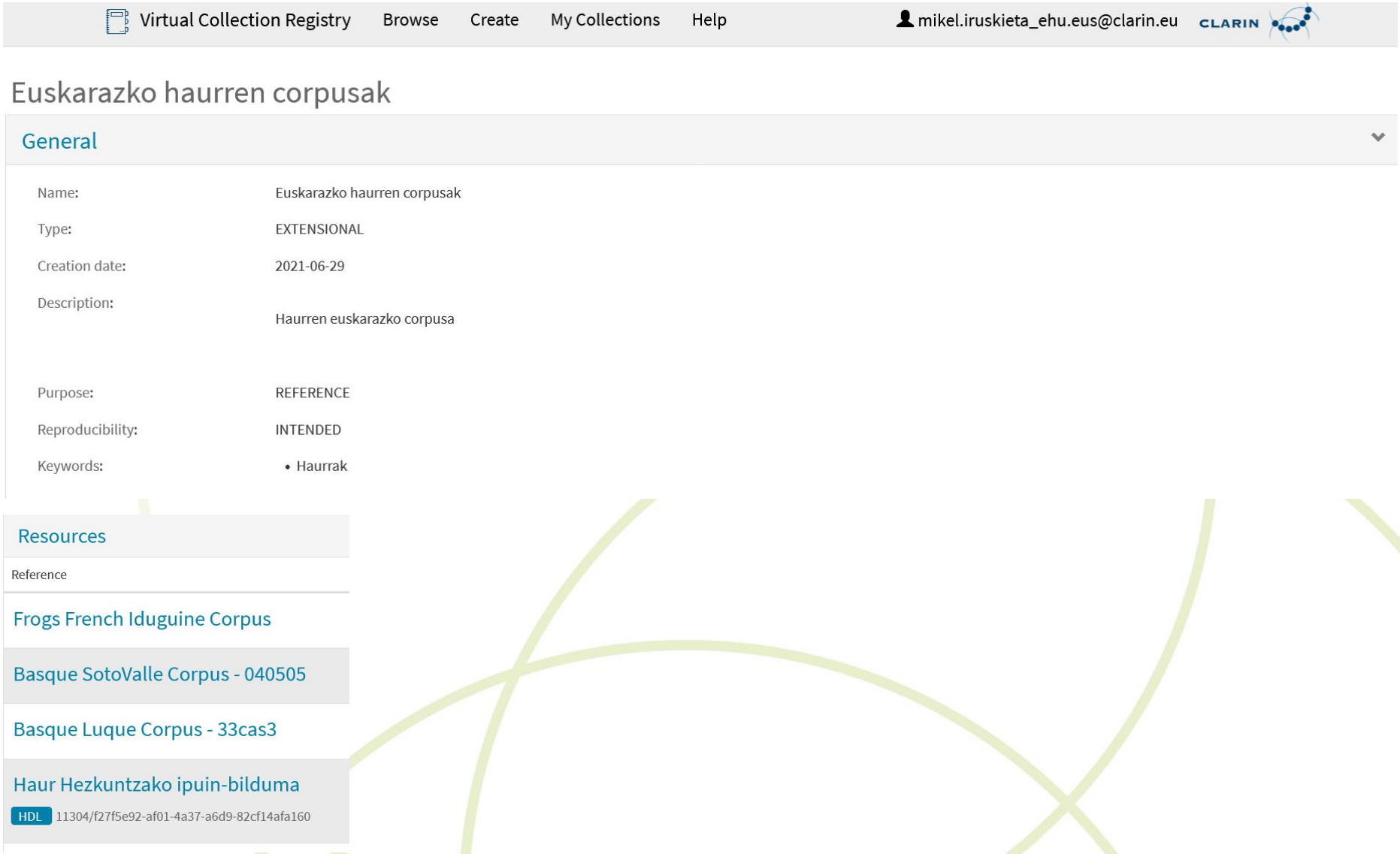
- Datuak:
  - “La edad de plata” por Jose Calvo Tello
  - Haurren corpusak
- Analisi sintaktikoa
- Analisia lainoan
- Argitalpen iraunkorrae



Interoperable

Originala ingelesez:
https://www.clarin.eu/showcase/eosc-portal-demonstration

CLARINeko bilduma birtual multimodala



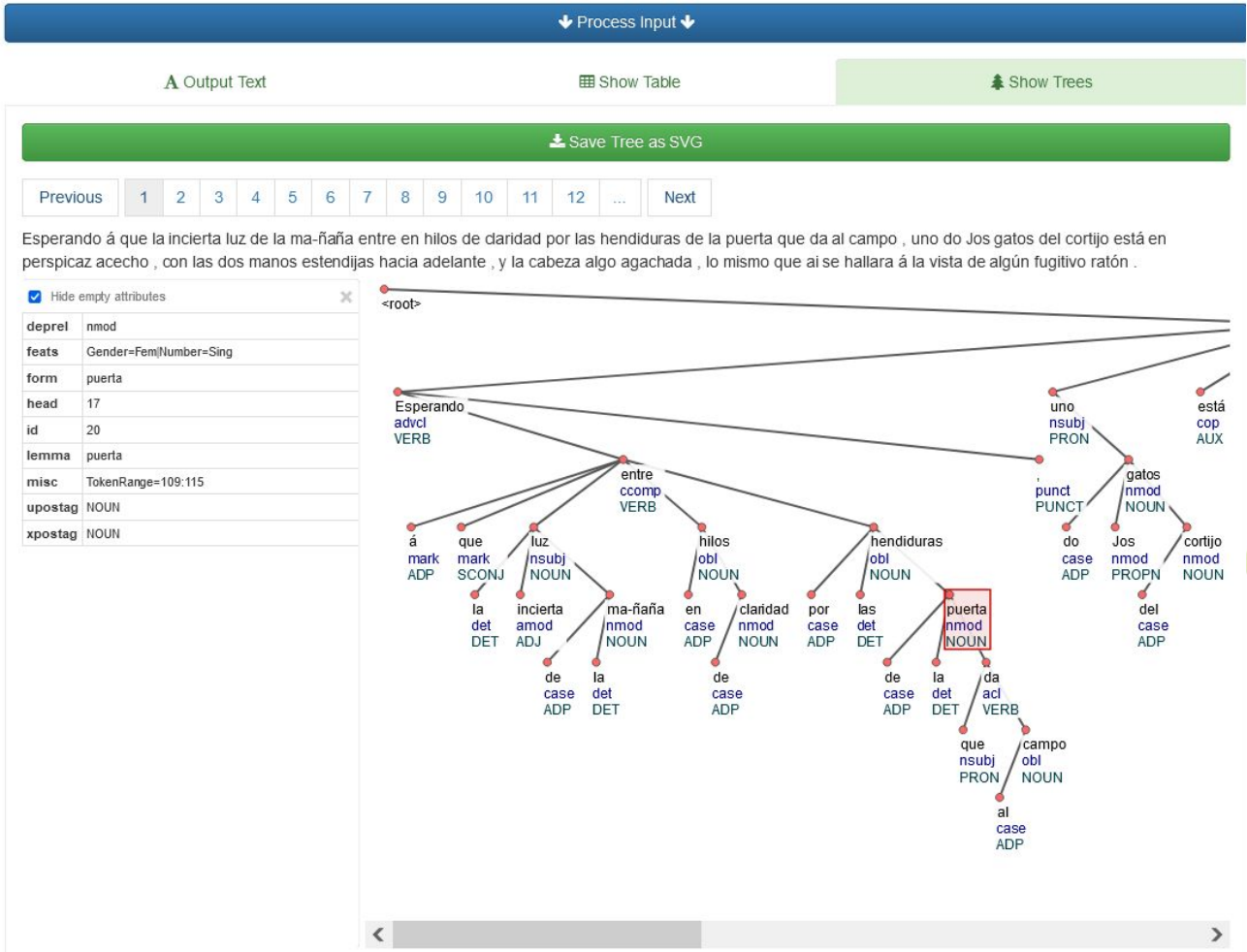
Idatzizko testurako baliabideak lainoan

1. Testua jaitsi: Gitzip
2. Gorde corpusa Eudaten
3. Aztertu Switchboardekin
4. Aukeratu tresna bat aurrera egieko
5. ...

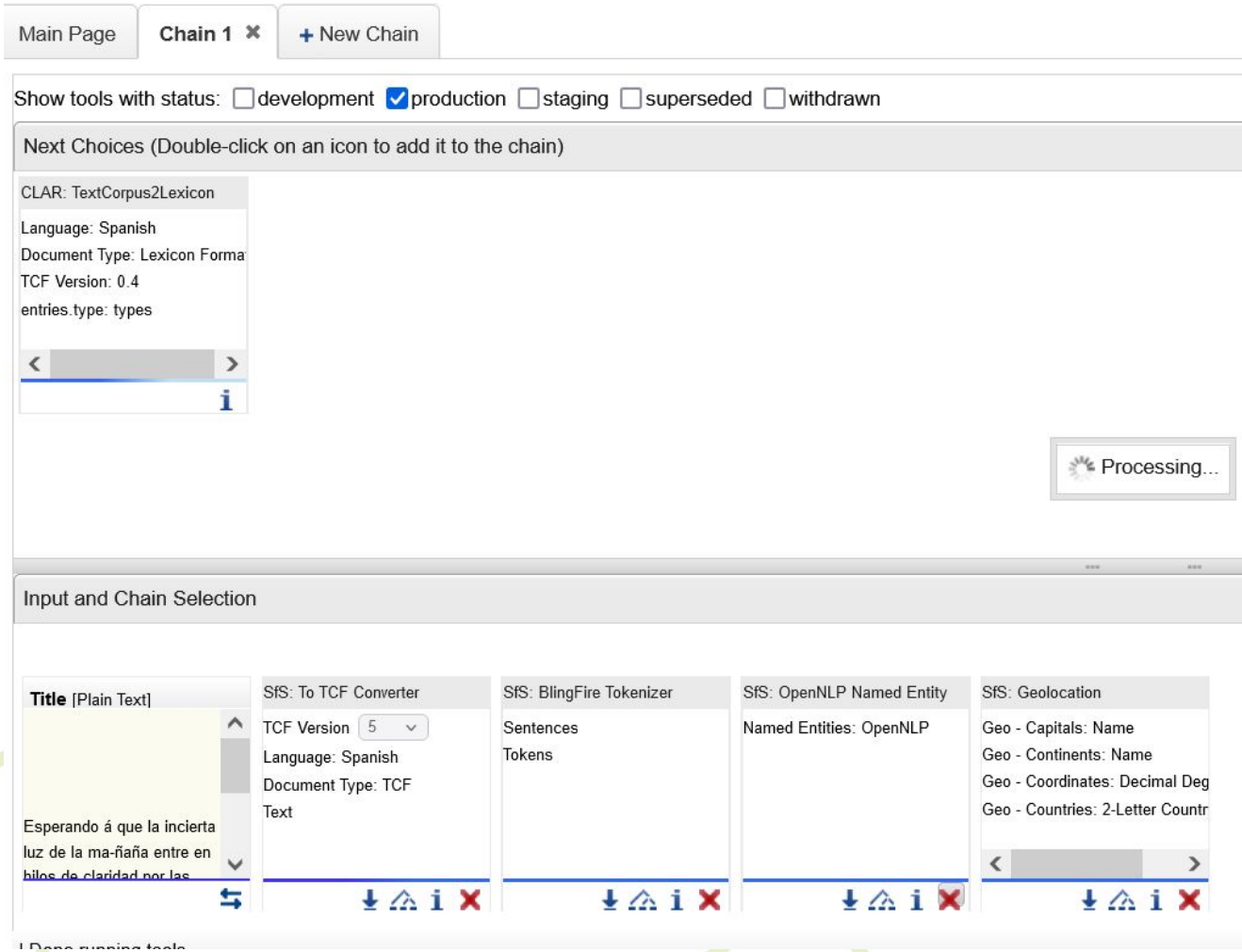
- Baliabideak
  - Euskararako: 3
  - Gaztelerarako: 4
  - Ingeleserako: 16
  - Alemanierarako: 13
  - Polonierarako: 26

1. Constituency Parsing
2. Coreference Resolution
3. Dependency Parsing
4. Distant Reading
5. Extraction of Polish terminology
6. Inclusion detection
7. Keyword Extractor
8. Lemmatization
9. Machine Translation
10. Metadata Processing
11. Morpho-syntactic tagger
12. Morphological Analysis
13. Named Entity Recognition
14. Named Entity Relation Detection
15. Part-Of-Speech Tagging
16. Sentiment Analysis
17. Shallow Parsing
18. Spatial expression detection
19. Speech Recognition
20. Stylometry
21. TF, IDF, TF-IDF calculation
22. Text Analytics
23. Text Enhancement
24. Text Summarization
25. Tokenisation
26. Topic Modelling
27. Visualisation of Geographic Data
28. Word sense disambiguation

Testu baten analisi sintaktikoa eta bistaratzea



Klik eginez, analisi-katea eraikiz



Distant Reading for European Literary

				AUTHORSHIP				LENGTH			TIME SLOT		
Language	Last update	Texts	Words	Male	Female	1-title	3-title	Short	Medium	Long	1840-59	1860-79	1880-99
cze	2021-04-09	100	5621667	88	12	62	6	43	49	8	12	21	39
deu	2021-04-11	100	12738842	67	33	35	9	20	37	43	25	25	25
eng	2021-04-09	100	12227703	49	51	70	10	27	27	46	21	22	31
fra	2021-04-09	100	8712219	66	34	58	10	32	38	30	25	25	25
gsw	2021-06-07	38	2392060	21	17	13	5	11	22	5	0	1	13
hrv	2021-03-22	21	1440018	21	0	4	0	6	12	3	6	12	2
hun	2021-04-09	100	6948590	79	21	71	9	47	31	22	22	21	27
ita	2019-11-21	34	3328244	32	2	19	3	13	10	11	5	12	10

INTELEko workshopa: "Distant Reading for European Literary History"

Saio praktikoa: github.com/bncolorado/Processing-ELTeC-corpus

TEI+XML fitxategiaren analisia ELTeC

Language Resource Switchboard | Upload | Tool Inventory | Help

Resources

SPA2013_OrtegaYFrias_ElDuende.xml 1.12 MiB

Mediatype: application/tei+xml

Matching Tools

Distant Reading

Open Voyant Tools

Voyant Tools

Cirrus | Terms | Links | Reader | TermsBerry | Trends | Document Terms

El duende de la corte : edición ELTeC(Ortega y Frías, Ramón (1825-1883))

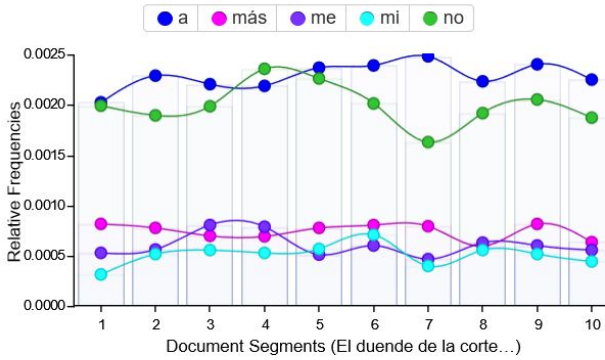
El duende de la corte

o

Memorias de un fraile

Novela histórica original

de



Terms:

Reset | Display

Summary | Documents | Phrases | Contexts | Bubblelines | Correlations

This corpus has 1 document with 178,052 total words and 14,073 unique word forms. Created now.

Vocabulary Density: 0.079

Average Words Per Sentence: 15.2

Most frequent words in the corpus: a (4049); no (3542); más (1316); me (1076); mi (909)

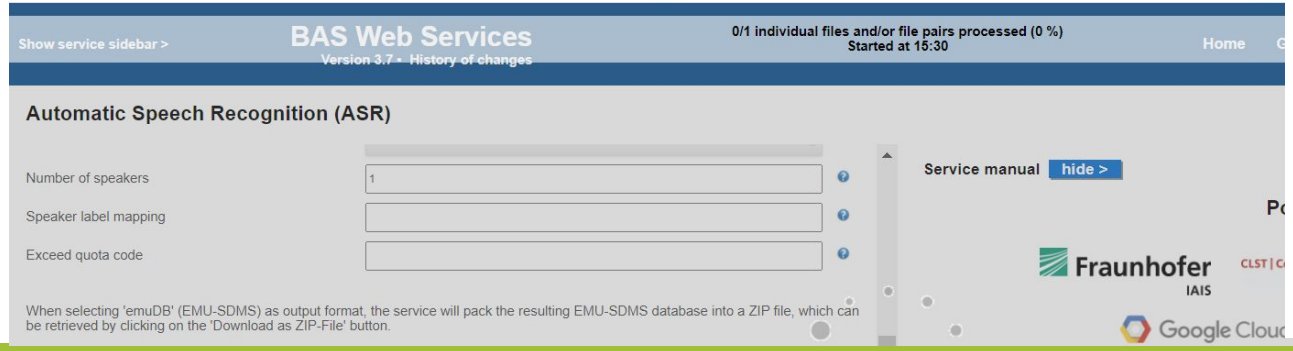
items:

Document	Left	Term	Right
1) El du…	fácilmente sin necesidad de pedirlos	a	la imaginación del poeta, cuyas
1) El du…	vez se arranca una lágrima	a	los ojos, un suspiro al
1) El du…	una historia lo que voy	a	referir. No he tenido que
1) El du…	lectura de algunos párrafos convenció	a	mi amigo de que había
1) El du…	de cedérmelo. [1] Así vino	a	mis manos esta historia, y

4,049 context expand

Ahotserako baliabideak lainoan

1. Deskargatu EJko bideoa
2. Transkribatu klik batekin BASetik
3. Emaitzak aztertu



1. Mary TTS
2. ASR
3. TextAlign
4. Pipeline without ASR
5. Pho2Syl
6. Chunker
7. AnnotConv
8. G2P
9. OCTRA - online text transcription system.
10. AudioEnhance
11. WebMAUS General
12. Chunk Preparation
13. Coala
14. WebMINNI
15. WebMAUS Basic
16. Anonymizer
17. TextEnhance
18. Formant Analysis
19. Subtitle
20. EMU Magic
21. Voice Activity Detection
22. EMU webApp - online labeling of speech data and more.
23. Pipeline with ASR
24. SpeakDiar

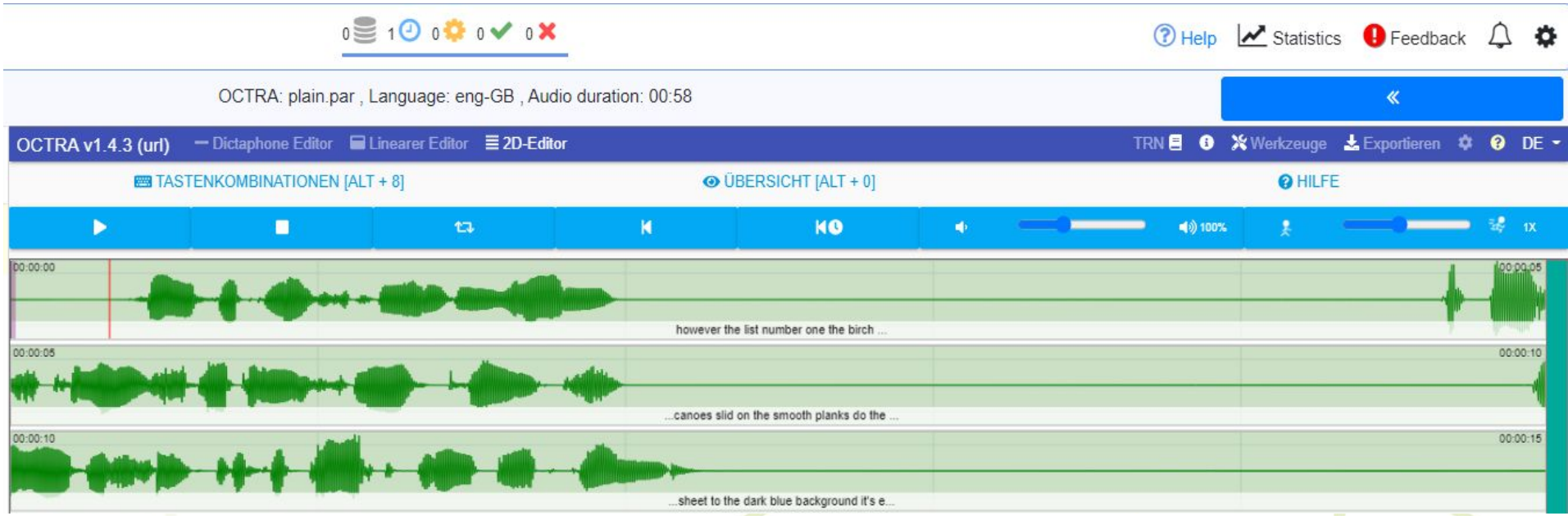
Euskera-gaztelania transkribapena

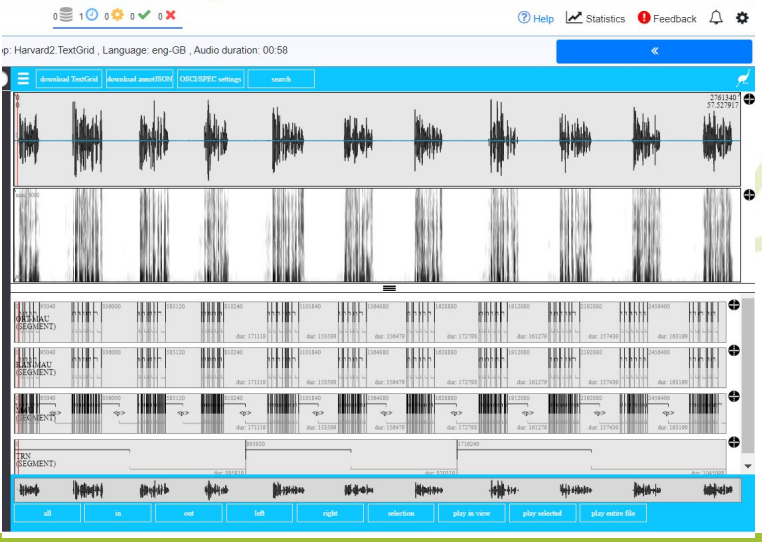
ITURRIA:
https://www.legebiltzarra.eus/portal/eu/web/eusko-legebiltzarra/noticias-y-eventos/actos-y-eventos/-/buscador/content/40-aniversario-del-parlamento-vasco-una-mirada-retrospectiva

callGoogleASR: egun on guztioi eta ongi etorri abestia eusko legebiltzarrak euskal herriko unibertsitatearen udako ikastaroen baitan antolatu duen 2000 goiko ikastaro honetara eskerrak eman nahi dizkizuet jardunaldi hauetan parte hartu duzuen guztioi hizlari partehartzaile antolatzaileei ere gehiago covid-19 da gure bizitzak etengabe baldintzatzen dituen une honetan ikastaro hau horren lekuko eusko legebiltzarraren 40. urteurrena atzerako begirada da ikasturte honetarako aukeratutako gaia ezin ziteken besterik izan izan ere aurten 40 (...) izan gara eta legebiltzarrak horretan paper garrantzitsua izan du en estos dos legislaturas el parlamento vasco se ha ido construyendo y consolidando dia a dia del mismo modo que este pueblo nuestro pueblo se ha ido reconstruyendo la trayectoria de la camara ha sido y es fiel reflejo de la evolucion social la presencia de la mujer (...) eta konpromisoz aurre egiteko zuen ekarpenak helburu horretan lagunduko dugula sinetsita berriz ere eskerrak eman nahi dizkizuet guztioi

OH Portal (CLARIN)

https://oralhistory.eu/oh-portal

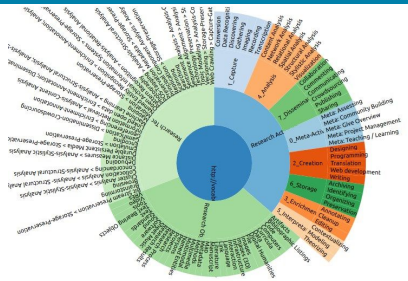




Webinar: https://youtu.be/X6bFGJpMjVQ

	Poloniera	Alemaniera	Inglesa	Gaztelania	Euskara
1. Constituency Parsing		x	x		
2. Coreference Resolution	x				
3. Dependency Parsing	x	x	x	x	x
4. Distant Reading	x	x	x	x	x
5. Extraction of Polish terminology	x				
6. Inclusion detection	x				
7. Keyword Extractor	x				
8. Lemmatization		x	x		
9. Machine Translation		x	x		
10. Metadata Processing					
11. Morpho-syntactic tagger	x		x		
12. Morphological Analysis	x	x	x		
13. Named Entity Recognition	x	x	x	x	
14. Named Entity Relation Detection					
15. Part-Of-Speech Tagging	x	x	x		
16. Sentiment Analysis	x				
17. Shallow Parsing	x				
18. Spatial expression detection	x				
19. Speech Recognition					

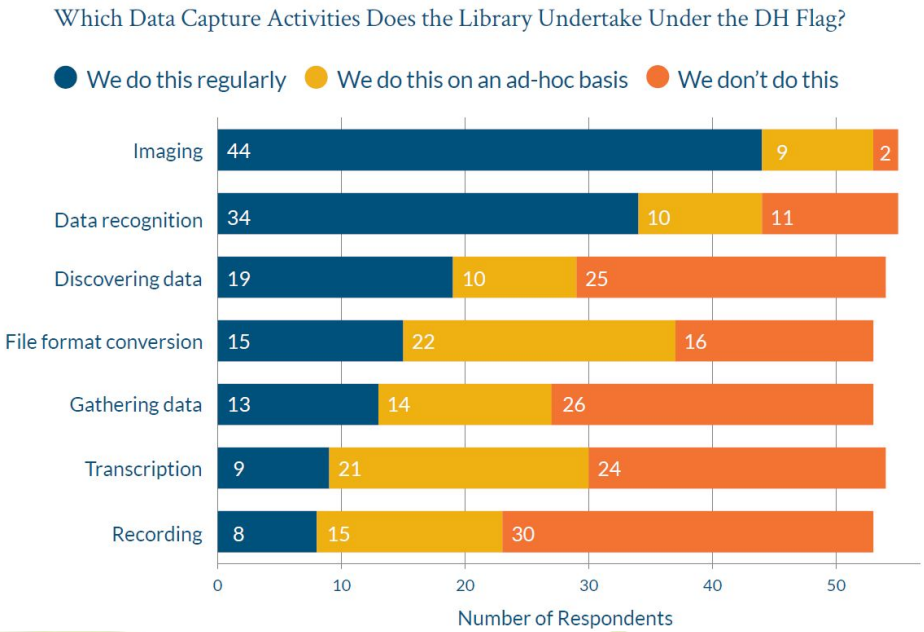
Azpiegiturak barne-egiturak sendotzeko



TADIRAH: Taxonomy

Zientzia irekia erreproduzigarria eta eraginkorra

- Lankidetzan aritu
- Ikerketak eta emaitzak plaza batean batzeko
- Bir-erabilpena bultzatzeko
- Diziplinartekotasuna sustatzeko
- Eragin sozioala handitzeko



FUENTE:
https://zenodo.org/record/3247286#.YV664RzRahM

Europe's Digital Humanities Landscape: A Study From LIBER's Digital Humanities & Digital Cultural Heritage Working Group

MERSI danke 謝謝 рахмат Баярлалаа спасибо faafetai lava kiitos dankie nandri nanni dhanyavad bedankt bayarlalaa gracie hvala mauruuru enkosi dziękuję sobodi dekuji obrigado mesi sagolun didi madloba najis tuke kam sah hamnida rahmat ESKERRIK ASKO sukriya kop khun krap ありがとう terima kasih 감사합니다 xiexie THANK YOU vinaka blagodaram mersi barka welalin tack dank je teşekkür ederim misaotra matondo paldies grazzi mahalo tapadh leat gracias хвала asante manana obrigada murakoze tenki djiere dieuf tau mochchakkeram дякую mamnun chokrane cnnorakaloutioun gratias ago gracies sulpay taiku go raibh maith agat arigatō takk dakujem trugarez tanemirt grazie rahmet dhanyavadagalu diolch shukriya merce мерси ευχαριστώ merci شكريا Schukria

Erreferentziak

Iruskieta, M. (2019) CLARIN Europako sarea: eHumanitateak eta zientzia sozialak lankidetzarako behar duten hizkuntza-azpiegitura sortzen. Humanitate digitalak: aukerak, erakundeen rol berriak eta elkarlana. UEU. 2019ko ekainaren 20a. Bilbo. ULR: https://www.youtube.com/watch?v=EEAwzxPL4GA&t=1346s

Krauwer, S., & Hinrichs, E. (2014). The CLARIN research infrastructure: resources and tools for e-humanities scholars. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC-2014)* (pp. 1525-1531). European Language Resources Association (ELRA).

Váradi, T., Wittenburg, P., Krauwer, S., Wynne, M., & Koskenniemi, K. (2008). CLARIN: Common language resources and technology infrastructure. In *6th International Conference on Language Resources and Evaluation (LREC 2008)*.

Bel, N. Gonzalez-Blanco, E. Iruskieta, M. (2016). CLARIN Centro-K-español. Procesamiento del Lenguaje Natural 57: 151-154. ISSN: 1135-5948.

CLARIN: https://www.clarin.eu/

DARIAH: https://www.dariah.eu/

INTELE: http://ixa2.si.ehu.eus/intele/