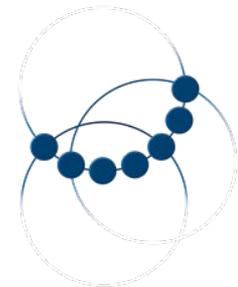# CKCMC: CLARIN K(nowledge)-Centre for CMC and Social Media Corpora

Egon W. Stemle, Jennifer-Carmen Frey, Alexander König, Achille Falaise, Tomaž Erjavec, Harald Lüngen

eurac research

CLARIN

Laboratoire de linguistique formelle

Jožef Stefan Institute

IDS | LEIBNIZ-INSTITUT FÜR DEUTSCHE SPRACHE

Welcome

CLARIN K CENTRE

[Contact]

CLARIN K(nowledge)-Centre for Computer-Mediated Communication and Social Media Corpora

The CLARIN Knowledge Centre for Computer-Mediated Communication and Social Media Corpora (CKCMC) offers expertise on language resources and technologies for Computer-Mediated Communication and Social Media. Its basic activities are to

1. Give researchers, students, and other interested parties information about the available resources, technologies, and community activities,
2. Support interested parties in producing, modifying or publishing relevant resources and technologies and
3. Organize training activities.

Computer-Mediated Communication (CMC)

User-generated CMC and social media content offers a wide range of research opportunities for a growing multidisciplinary research community to examine themes that often relate to—but are not limited to—the interaction between language, CMC, and society like, for example, language variation, pragmatics, media and communication studies. The data is also very important for the development

Welcome

CLARIN K CENTRE

Contact

CLARIN K(nowledge)-Centre for Computer-Mediated Communication and Social Media Corpora

The CLARIN Knowledge Centre for Computer-Mediated Communication and Social Media Corpora (CKCMC) offers expertise on language resources and technologies for Computer-Mediated Communication and Social Media. Its basic activities are to

1. Give researchers, students, and other interested parties information about the available resources, technologies, and community activities,
2. Support interested parties in producing, modifying or publishing relevant resources and technologies and
3. Organize training activities.

Computer-Mediated Communication (CMC)

User-generated CMC and social media content offers a wide range of research opportunities for a growing multidisciplinary research community to examine themes that often relate to—but are not limited to—the interaction between language, CMC, and society like, for example, language variation, pragmatics, media and communi-cation studies. The data is also very important for the development

history of the German language.

Contact

Our helpdesk can be contacted via email to `helpdesk @ THIS DOMAIN`. The helpdesk offers additional clarifications regarding the documentation and support in using, modifying, producing, or publishing CMC resources and technologies.

Get in touch!   Check out previous inquiries…›

Documentation

For more detailed information, see the dedicated documentation section.

Conference Series ▾   CLARIN K-Centre for CMC ▾   CMC-TEI   Publications   News   Contact   Q

Q Search...   /

CKCMC Docs

FAIR-ify

CMC-TEI

CLARIN K(nowledge)-Centre for CMC Documentation

We offer expertise on language resources and technologies for Computer-Mediated Communication and Social Media Corpora. We can be contacted via email to `helpdesk @ THIS DOMAIN`. There, we offer additional clarifications regarding this documentation and support in using, modifying, producing, or publishing CMC resources and technologies. For more general questions, also consider joining the European group of researchers interested in the development and processing of CMC corpora @ Google Groups.

As we are a young K-Centre, our documentation is still under development. Once we have collected enough input, we will be adding a FAQ section with your most pressing questions. We are also interested in your thoughts on missing or wrong information.

In the meantime, we have already started to summarise important information for the two central topics of our K-Centre:
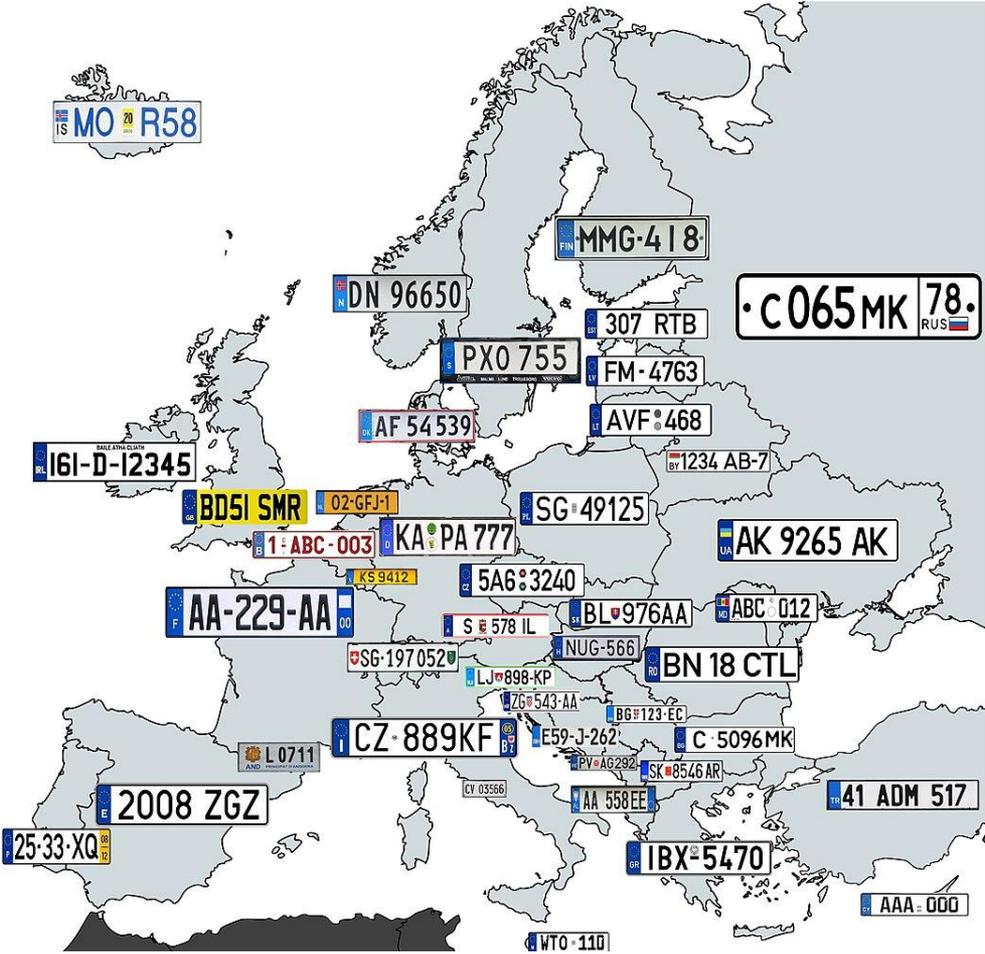
- FAIR best practices when creating a new CMC corpus
- CMC-TEI about a TEI standard for the representation of the structural and linguistic peculiarities of CMC

F-indability A I R Principles

F-indability A I R Principles

F A-ccessibility I R Principles

F A-ccessibility I R Principles

F A I-nteroperability R Principles

F A I-nteroperability R Principles

F A I R-euse Principles

Home / About / CLARIN Centres / Overview CLARIN Centres

Overview CLARIN Centres

All currently established CLARIN centres are listed below. This list is based on the Centre Registry, a database that is the authoritative source for information about each of the centres. If you would like to become a CLARIN centre, please fill out this form.

Centre	Type	City	Consortium
Austrian Centre for Digital Humanities and Cultural Heritage - A Resource Centre for the HumanitiEs	B	Vienna	Digital Humanities Austria (AT)
Bayerisches Archiv für Sprachsignale	B	München	CLARIN-D (DE)
Berlin-Brandenburg Academy of Sciences and Humanities	B	Berlin	CLARIN-D (DE)
Center of Estonian Language Resources (CELR-EKK)	B	Tartu	CELR (EE)
Centre for Language and Speech Technology	C	Nijmegen	CLARIAH (NL)
Centre for the Digital Foundation of Research in the Humanities, Social, and Educational Sciences	C	Frankfurt, Germany	CLARIN-D (DE)
CLARIN Centre of Latvian language resources and tools	C	Riga	CLARIN-LV (LV)

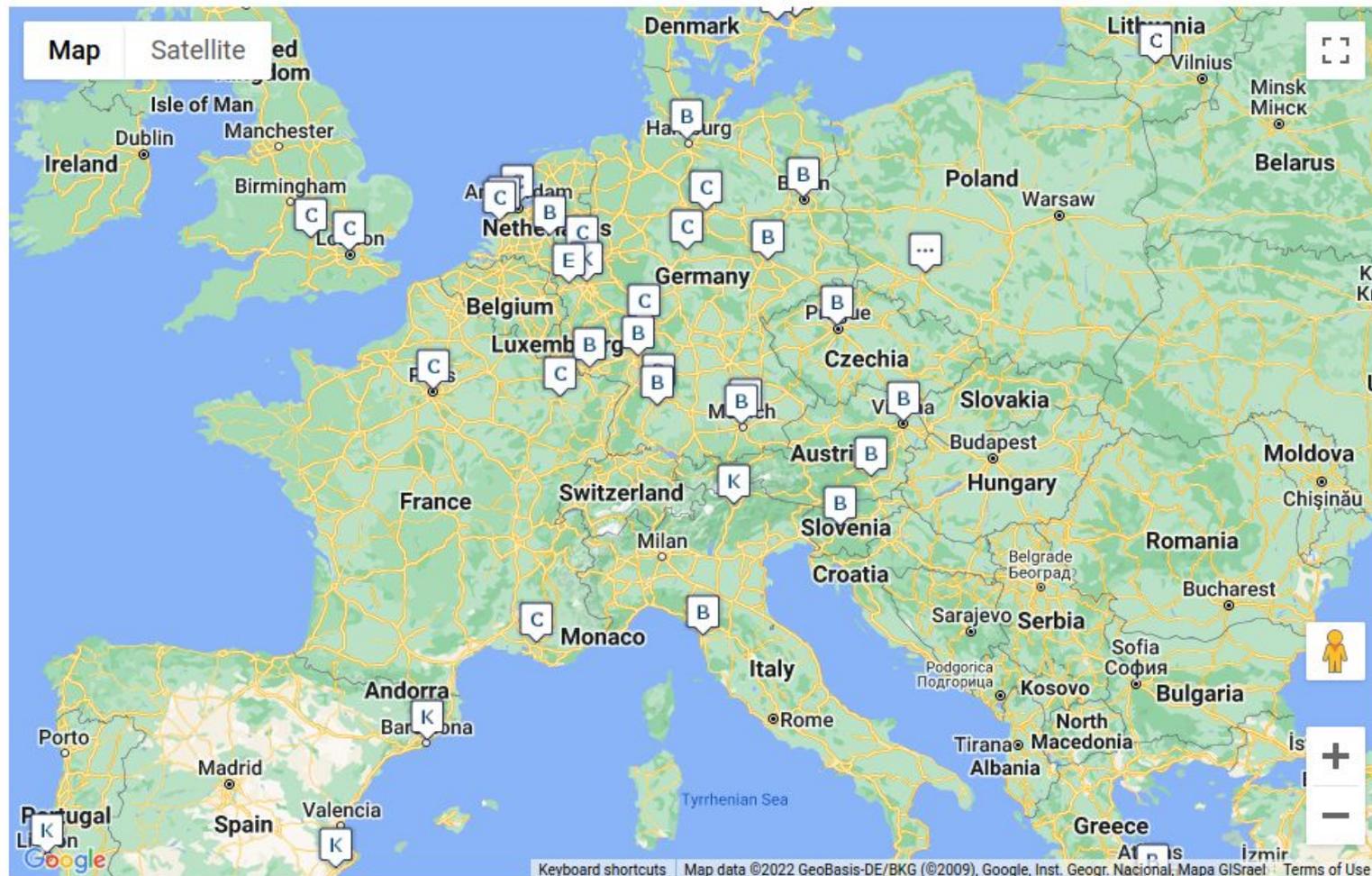ZIM Centre for Information Modelling   B   Graz   CLARIAH-AT (AT)

🌐 Virtual Language Observatory   **Search**   Contributors   Help   **CLARIN** ⦿⦿⦿⦿

VLO / Faceted search / Search results   ◁▾ 🏳

computer-mediated communication   🔍

Showing 1 to 10 of 72 results for (computer-mediated communication ✕) ❶   **Results per page:** [10 ⌄]

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

| Language | ⌄ |

| Collection | ⌄ |

| Resource type | ⌄ |

| Modality | ⌄ |

| Format | ⌄ |

⌄

| Temporal Coverage | ⌄ |

| Availability | ⌄ |

| Search options | ⌄ |

<< < **1** 2 3 4 5 6 7 8 > >>

CMC training corpus Janes-Tag 1.2   3📄 3🗋
(Part of CLARIN.SI data & tools)

⊞ Janes-Tag is a manually annotated corpus of Slovene **Computer-Mediated Communication** (CMC). It is meant as a gold-standard training and testing dataset for tokenisation, sentence segmentation, word normalisation, morphosyntactic tagging and lemmatisation of non-standard Slovene. As the corpus has been carefully manually…   ©️ ℗ ⓘ ⊜

(Slovenian)

🏠 Landing page for this record

CMC training corpus Janes-Tag 1.0   3📄 2🗋
(Part of CLARIN.SI data & tools)

⊞ Janes-Tag is a manually annotated corpus of Slovene **Computer-Mediated Communication** (CMC). It is meant as a gold-standard training and testing dataset for tokenisation, sentence segmentation, word normalisation, morphosyntactic tagging and lemmatisation of non-standard Slovene. As the corpus has been carefully manually…   ©️ ℗ ⓘ ⊜

(Slovenian)

🏠 Landing page for this record

Home   **Repository**   **About**   CLARIN 🔵

→] **Login**

Eurac Research CLARIN Centre Repository Home / View Item

Search 🔍

DIDI - The DiDi Corpus of South Tyrolean CMC 1.0.0   🖨

ERCC eurac research

> ❝ **Please use the following text to cite this item or export to a predefined format:** BIBTEX CMDI
>
> Frey, Jennifer-Carmen; Glaznieks, Aivars and Stemle, Egon W., 2019, *DIDI - The DiDi Corpus of South Tyrolean CMC 1.0.0*, Eurac Research CLARIN Centre, **http://hdl.handle.net/20.500.12124/7.**   📋

◉ **Browse**

> All of the Repository ▼

❮ Share: 📘 🐦 🔵

👤 **My Account**

→] Login

CMC & WaC

📊 **Statistics**

🖋 **Authors**	Frey, Jennifer-Carmen ; Glaznieks, Aivars ; Stemle, Egon W.
➴ **Item identifier**	http://hdl.handle.net/20.500.12124/7 📋
⚲ **Project URL**	http://www.eurac.edu/didi
⌇ **Demo URL**	https://commul.eurac.edu/annis/didi
⚲ **Is Based On**	https://gitlab.inf.unibz.it/commul/didi/data-bundle/-/tags/v1.0.0
⚲ **Referenced by**	http://www.eurac.edu/en/research/autonomies/commul/Documents/DiDi/NLP4CMC-2015_DiDi_paper.pdf http://www.eurac.edu/en/research/autonomies/commul/Documents/DiDi/didi_clic-it2016_FINAL.pdf
📅 **Date issued**	2019-03-07
🏷 **Type**	corpus
✖ **Size**	23000 texts, 370000 tokens

📉 Statistics   BETA

ⓘ **General Information**

↥ Deposit

❞ Cite

🔃 Submission Lifecycle

? FAQ

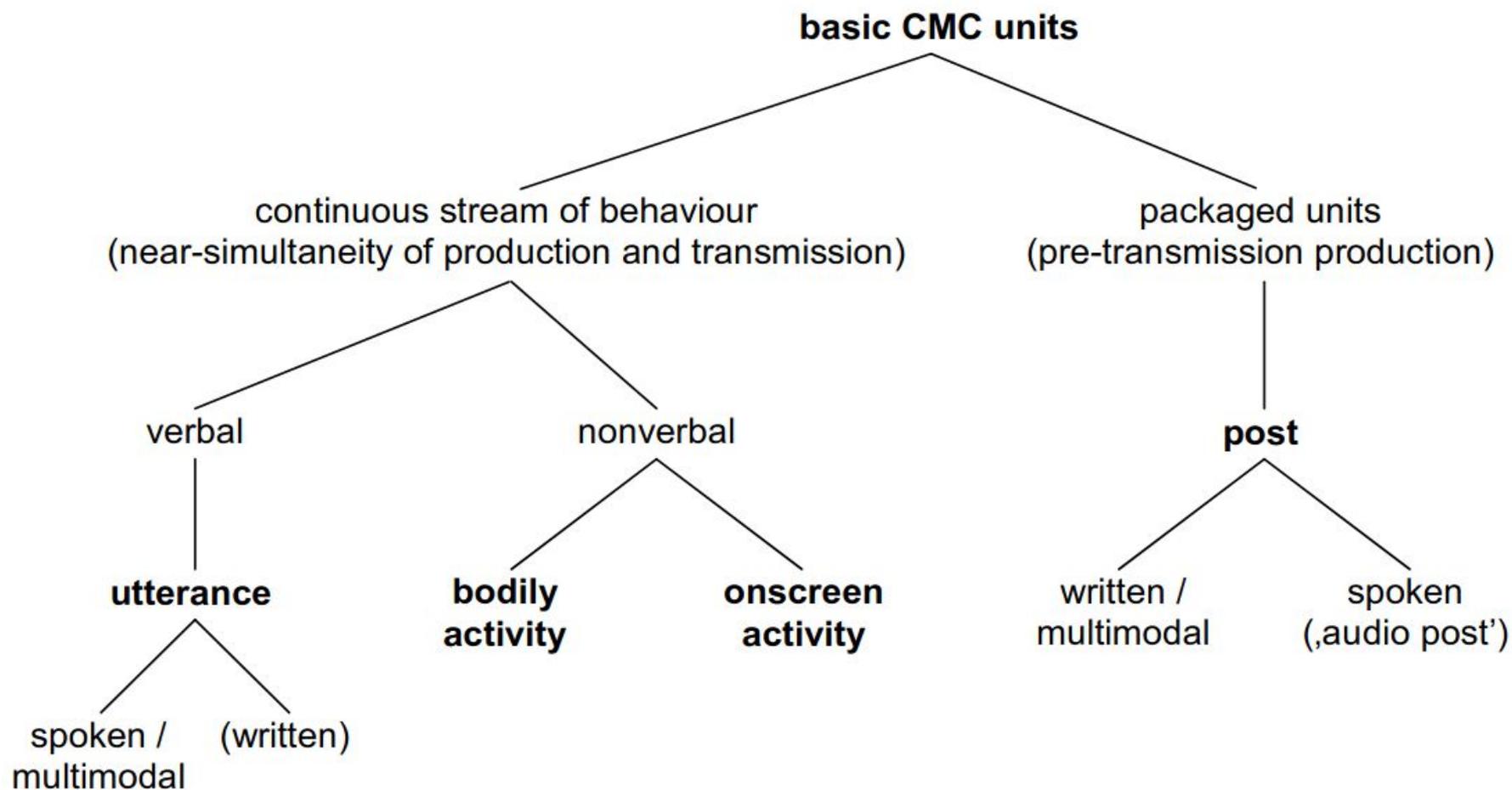❶ About

✉ Help Desk

Text Encoding Initiative (TEI)

The TEI consortium collectively develops and maintains a standard for the representation of texts in digital form.

Since 1994, the TEI Guidelines have been widely used by libraries, museums, publishers, and individual scholars to present texts for online research, teaching, and preservation.

Text Encoding Initiative (TEI)

Text Encoding Initiative (TEI) CMC-core

Questions

- Datasets used for today's presentations online?
- Metadata available?
- Corpus paper available (detailed provenance)?

https://cmc-corpora/ckcmc