

[Click here to be redirected to the virtual room of the Project Expo.](#)

**Project abbreviation:** E3C

**Project name:** European Clinical Case Corpus

**Project coordinator:** Bernardo Magnini

**Project consortium:** Bruno Kessler Foundation (FBK)

**Funding:** ELG (European Language Grid) Pilot Projects Open Call 1 (Grant Agreement No. 825627 – H2020, ICT 2018-2020 FSTP) 137K Euro

**Project duration:** 1 year

**Main key words:** information extraction for the medical domain; clinical entities tagging; temporal relations.

**Background of the research topic:** E3C builds on top of research on information extraction from medical texts. The development of technologies in this field suffers from the limited availability of annotated training data, which are necessary to feed deep learning algorithms. This situation is mainly due both to privacy issues on medical documents (e.g., hospital discharge summaries) and to the complexity of domain annotations in specialized domains.

**Goal of the project:** E3C aims to collect and annotate a multilingual corpus of clinical narratives, aspiring to become a reference European resource.

**Project abstract:** E3C aims to collect and annotate a multilingual corpus of clinical narratives, aspiring to become a reference European resource. A clinical narrative is a statement of a clinical practice, presenting the reason for a clinical visit, the description of physical exams, and the assessment of the patient's situation. We focus on published clinical narratives because they are often de-identified, overcoming privacy issues, and are rich in clinical entities as well as temporal information, which are almost absent in other clinical documents (e.g., radiological reports). E3C has built a 5-language (Italian, English, Spanish, French and Basque) clinical narrative corpus to allow linguistic analysis, benchmarking, and training of information extraction systems. The corpus includes two types of annotations: (i) *clinical entities*, i.e., disorders according to the UMLS clinical taxonomy; (ii) *temporal information*, i.e., events, time expressions and temporal relations, according to the THYME TimeML standard.

#### **Publications:**

Bernardo Magnini, Begona Altuna, Alberto Lavelli, Manuela Speranza, Roberto Zanolli. The E3C Project: European Clinical Case Corpus, proceedings of SEPLN 2021.

Bernardo Magnini, Begoña Altuna, Alberto Lavelli, Manuela Speranza, Roberto Zanolli. The E3C Project: Collection and Annotation of a Multilingual Corpus of Clinical Cases" proceedings of CLiC-it 2020.

