Use cases

### **Case 1 – European research project on spoken English corpus**

**Context:** A senior researcher is participating in a Horizon Europe project aimed at collecting and analyzing spontaneous spoken English from multilingual communities across Europe.

**DMP Objective:** To prepare an initial data management plan describing the types of data collected, standards for annotation and metadata, ethical and legal considerations, and long-term preservation and sharing strategies.

### **Case 2 – PhD in dialectology**

**Context:** A PhD student at the University of Naples “L’Orientale” is preparing a thesis project involving the construction of a digital corpus of oral and written texts in contemporary Neapolitan dialect for morphological and sociolinguistic analysis.

**DMP Objective:** To describe data collection, storage, and annotation procedures, address legal and ethical aspects related to personal data, and define conditions for long-term storage and controlled access.

### **Case 3 – Phonetic study on bilingual child speech**

**Context:** A research team in language acquisition is conducting a study on phonetic development in bilingual (Italian-English) children aged 3 to 6, using in-home audio recordings and phonetic transcription.

**DMP Objective:** To document procedures for data collection, storage, and annotation, address ethical issues related to sensitive data, and define conditions for long-term preservation and controlled reuse of the corpus.

### **Case 4 – Bilingual technical corpus (IT-EN)**

**Context:** A collaborative project between a university and translation professionals aims to build a bilingual aligned corpus of Italian-English technical manuals for machine translation research and development.

**DMP Objective:** To define data formats and alignment standards, describe metadata and legal usage agreements, and plan secure storage and public access with licensing for non-commercial reuse.

### **Case 5 – Diachronic study of legal English**

**Context:** A researcher is conducting a diachronic linguistic study of English legal texts spanning the 19th to 21st centuries, using digitized documents from online archives and libraries.

**DMP Objective:** To outline the data sourcing and cleaning process, clarify legal considerations for copyrighted materials, and define long-term archiving and open-access strategies.

### **Case 6 – Dataset for sentiment analysis**

**Context:** A university research group is developing a sentiment-labeled dataset of English product reviews to train and evaluate natural language processing models.

**DMP Objective:** To describe the methods for collecting and annotating review texts, define privacy protections, and provide a clear plan for public sharing of the dataset under an open license.

### **Case 7 – Clinical corpus in English for NLP**

**Context:** A research team at a national institute and hospital is building a pseudonymized corpus of English clinical reports to support the training of NLP models for medical applications.

**DMP Objective:** To address the secure handling of sensitive health data, document ethical approvals and anonymization processes, and outline long-term preservation and access under restricted licensing.