Example DMP

### 

**Context:** A PhD student must present to her committee a project that involves creating an annotated corpus of social media posts in English L2.

**DMP Objective:** Prepare a section dedicated to data management in the thesis proposal, explaining how the data will be collected (scraping, crowdsourcing), how legal issues will be addressed (licenses, copyright), and where it will be stored (institutional repository)

### 

### **Use Case – DMP schema completed**

**Project Title:** English Learner Language on Social Media  
**Project Manager:** Laura Bianchi, PhD student, Sapienza Università di Roma  
**Project Duration:** 01/11/2023 - 31/10/2026  
**Funder:** MIUR (PhD scolarship)  
**Institution Involved:** Sapienza Università di Roma

**Objectives:** To analyze the use of English L2 on social media by Italian learners, creating an annotated corpus for error analysis and interlanguage studies.

**Data:** New corpus; texts collected from social media, syntactic annotations and errors  
**Formats:** TXT, JSON, CoNLL-U, CSV  
**Volume:** ca. 10.000 post (1GB)  
**Collection:** Web scraping and semi-automatic manual annotation

**Ethics:** Potentially personal data; full anonymization; Terms of Use compliance

**Legal:** Check licenses and copyright

**Standards:** Universal Dependencies  
**Metadata:** Textual ID, type of error, estimated proficiency level of speakers  
**Tools:** Custom Python scripts, UDPipe

**Storage:** University cloud space with automatic backup

**Preservation:** Zenodo; Dublin Core metadata schema  
**Access:** Public on demand with authentication  
**Licensing:** CC BY 4.0

**Responsibility:** Laura Bianchi, PhD student  
**Retention:** 10 years  
**Sustainability:** Dataset versioning, GitHub documentation