# Template Data Management Plan (DMP) for a specific research output: Language Corpus

## 1. Project Description

* **Project Title:** [Insert project title]
* **Project Manager:** [Name, institution, contact]
* **Project Duration:** [Start date - End date]
* **Funder (if applicable)**: [Funding institution]
* **Institution Involved**:
* **Project Objectives:** [Describe briefly the objectives of the research and the role of the corpus]

## 2. Type and format of data

* Identification of the dataset
  + Are you re-using an already existing dataset?
  + Are you creating a new dataset?
  + Dataset name
* Types of data collected or generated:
  + Transcribed Texts
  + Audio/video files
  + Linguistic annotations
  + Speaker metadata / documents
  + Processed / tabular /structured / quantitative / qualitative
* Data formats:
  + Text: TXT, JSON, XML, TEI
  + Audio: WAV, MP3, FLAC
  + Annotations: ELAN (EAF), Praat (TextGrid), CSV
  + Metadata: CSV, XML
* How is the data generated / collected
* Estimated data volume:
  + [Give approximate amount of data expected, e.g. 100 hours of speech, 1TB of data, etc.]

## 3. Ethics and legal aspects

* Are you handling personal data?
  + Solutions adopted for handling personal data
* Legal aspects
  + Are there anyl egal issues associated to your output?
  + Are there other ethical issues associated with the dataset?

## 4. Standards and metadata

* **Annotation and transcription standards adopted:** 
  + [Es. TEI, ISO 24624, Universal Dependencies (UD), CMDI]
* **Metadata collected:** 
  + Unique document / Speaker ID
  + Region, age, gender of speakers (anonymised)
  + Communicative context (formal/informal)
  + Technical information on recordings
* **Metadata management tools:** 
  + [Es. ELAN, EXMARaLDA, OpenRefine, proprietary software]

5. Data storage and security

* **Where will data be stored during the project?** 
  + [Es. institution's secure server, internal repository]
* **Backup and security measures:** 
  + Automatic daily/weekly backups
  + Encryption for sensitive data
  + Access control with authentication

## 5. Making research FAIR

* **Long-term preservation**
  + Selecting what to preserve
  + Chosen repository/ies
* **Findability**
  + Persistent Identifiers
  + Metadata
  + Keywords
* **Data access level:** 
  + Audio/video files: [Restricted access / public with restrictions]
  + Audio/video files: [Restricted access / public with restrictions]
  + Annotations and metadata: [Public / accessible with restrictions]
  + Is metadata openly accessible?
* **Interoperability**
  + Methodologies for data generatione and processing
  + Vocabularies, tazonomies and other standards
* **Reusability**
  + Licensing [Es. CC BY 4.0, CC BY-NC 4.0, CC0, sccess by agreement]
  + Documentation

## 6. Responsibility and plan for sustainability

* **Responsible for data management:** 
  + [Name, role, contact]
* **Duration of data retention:** 
  + [Es. Minimum 10 years]
* **Plan for sustainability:** 
  + Use of open formats to ensure long-term accessibility
  + Periodic review of file integrity
  + Detailed documentation for re-use by the community

## Additional notes

[Insert any additional information or project-specific details]