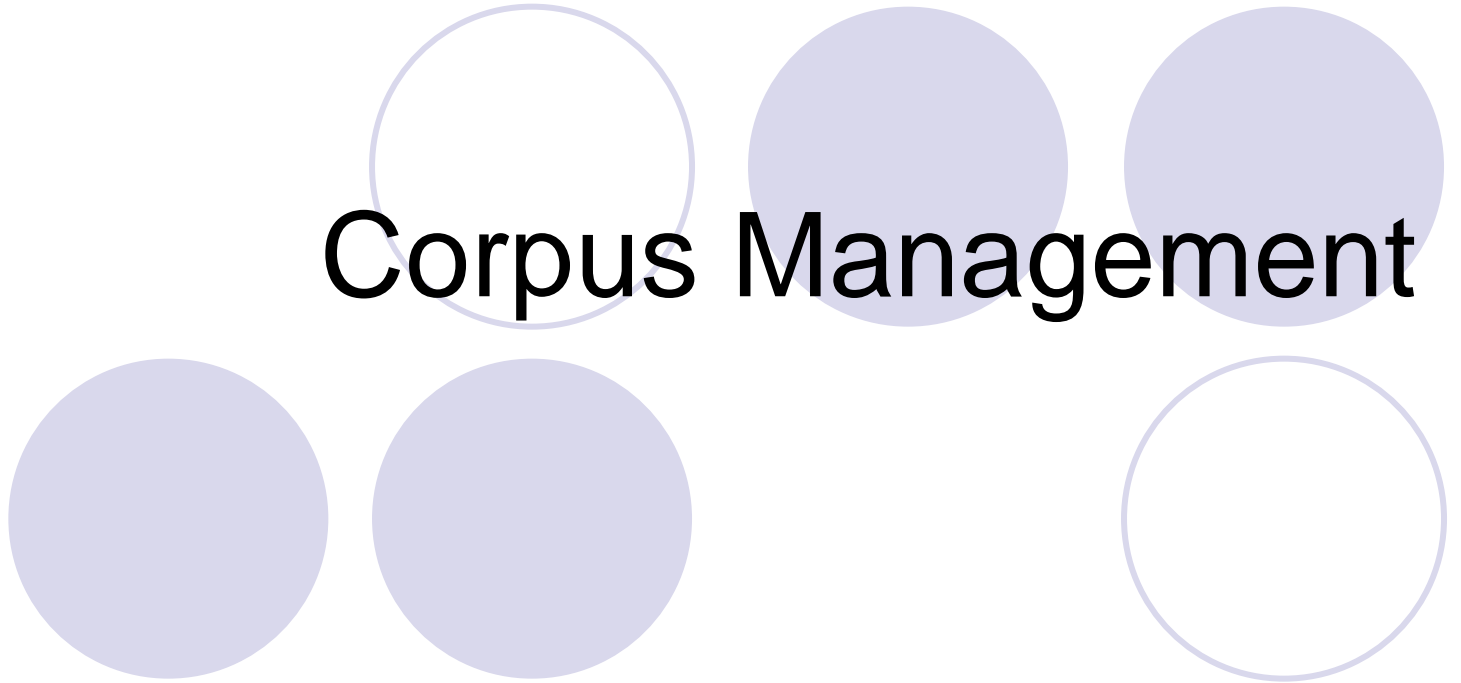


Corpus Management

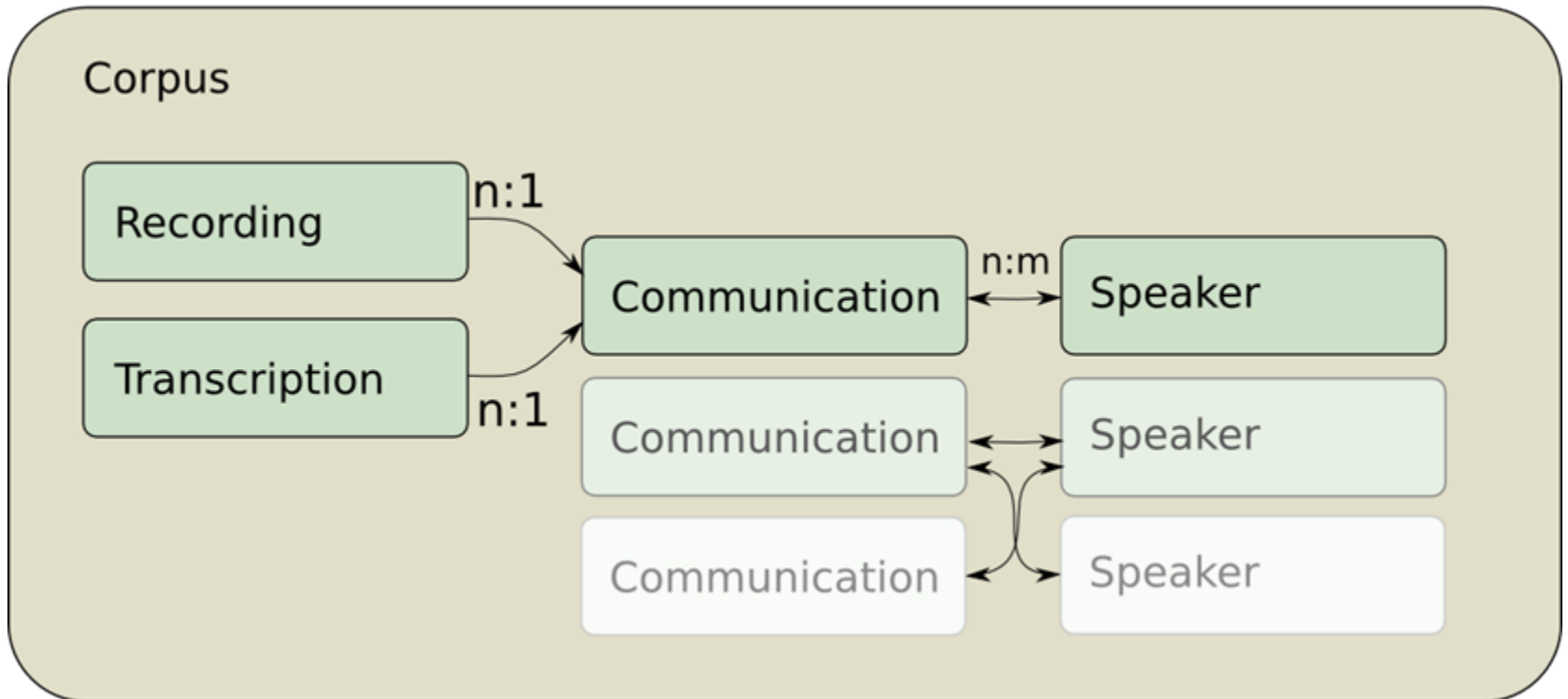




Corpus

- Collection of
 - audio and/or video recordings
 - their transcriptions
- + Information („Metadata“) about communications and speakers, such as
 - conversation type (e.g.: telephone, face-to-face, TV talkshow)
 - age, sex, place of birth, etc.

Corpus

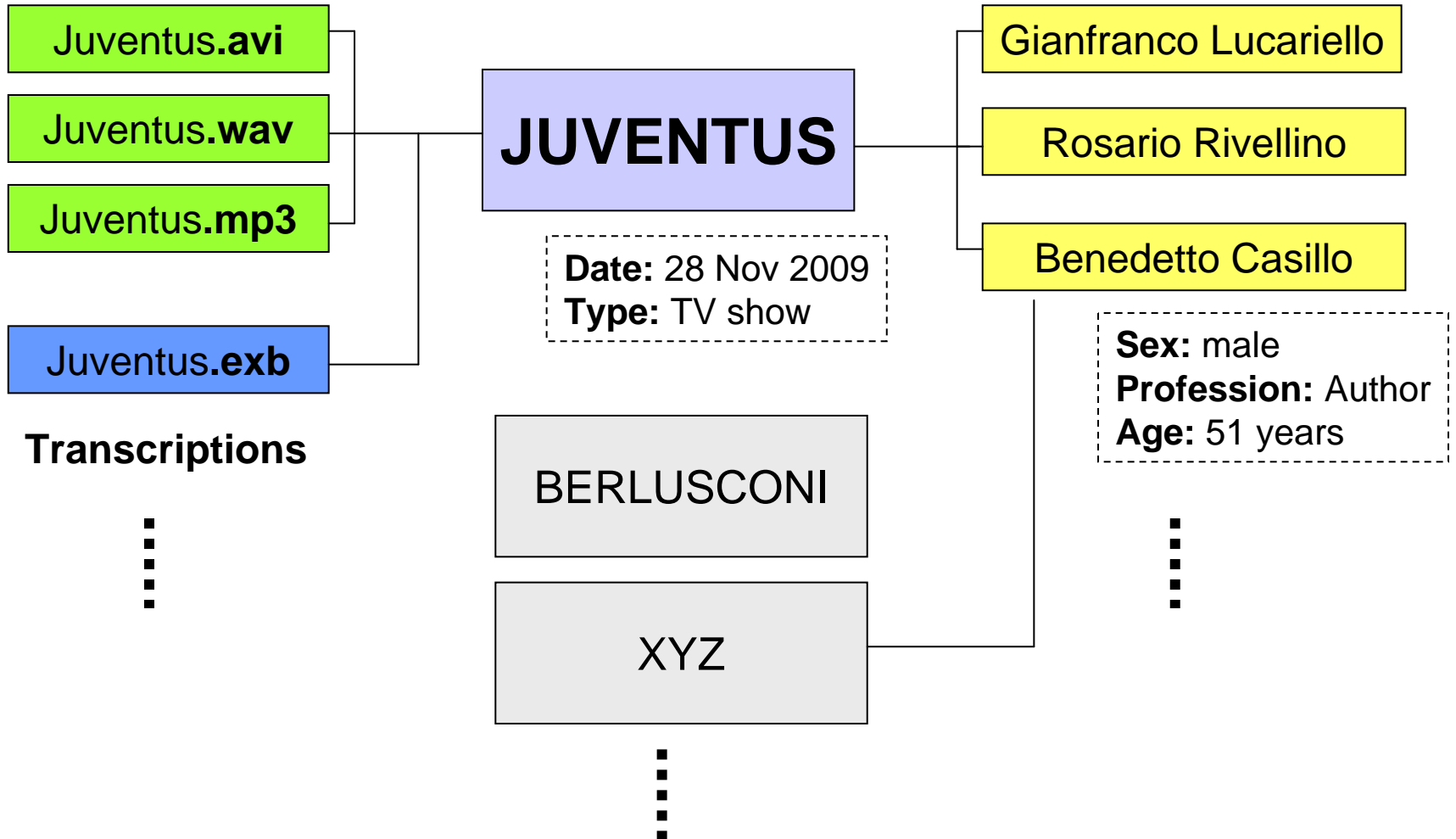


Corpus

Recordings

Communications

Speakers



Meta data



- ... to make a corpus understandable/reusable for others
- ... to relate linguistic data (transcription) and extralinguistic data in analysis, e.g.
 - „The phonetic influence of Spanish on Catalan will vary according to age of speakers and the proportion of Spanish speakers in the part of Barcelona they live in“
 - „Older Faroese speakers will use fewer Danish loanwords than Younger Faroese speakers“
 - „German Turks will use „German word order“ in their Turkish more frequently than Turks in Turkey“