



Physical textual documents

Transcribed and annotated files  
(.txt, XML, spreadsheets, databases, etc.)

Network structure

**Textual Sources  
Acquisition**

**Digitization**

**Annotation**

**Network  
Creation**

**Network  
Visualization and  
Analysis**

**P1:** Too many missing documents (data not exhaustive)

**P2:** Original sources and digitization mismatch (OCR error, transcription error, etc.)

**P3:** Missing annotation  
**P4:** Inconsistent annotations  
**P5:** Errors in Named Entity Recognition and Disambiguation

**P6:** Wrong network model

**P7:** Wrong representation  
**P8:** Wrong choice of algorithm/measure