

Between the folds

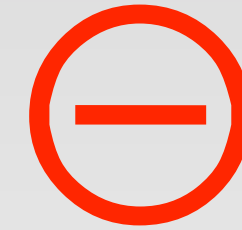
A hybrid model
for online publishing projects
integrating **philology**,
mass document repositories,
and automatic **text analysis**

We are not scanning all those books to be read **by people.**

We are scanning them to be read **by an AI.**

Google Books

mass digitization



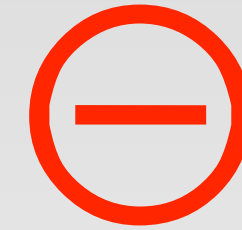
- massive amount of data
- automatic extraction of metadata

- inaccurate data
- restricted metadata
- undocumented editing procedure

TEI-based editing projects



- accurate data
- extensive metadata
- documented editing procedure



- restricted amount of data
- automatic text analysis is less feasible

CORE

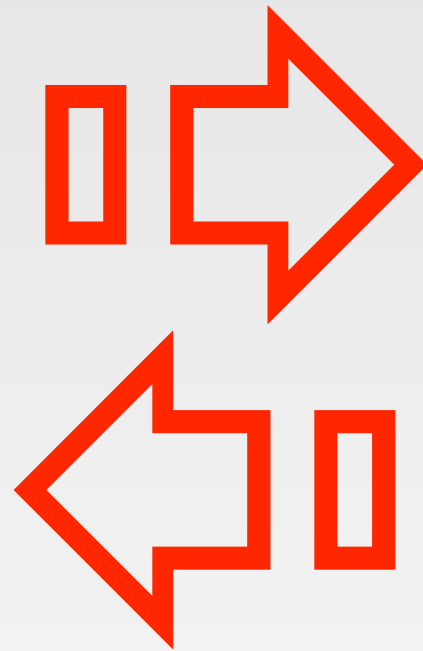
INTERFACE

CLOUD

TEI project



web
interface



APIs



mass document
repositories

Proof of concept

theater.ua.ac.be/cti