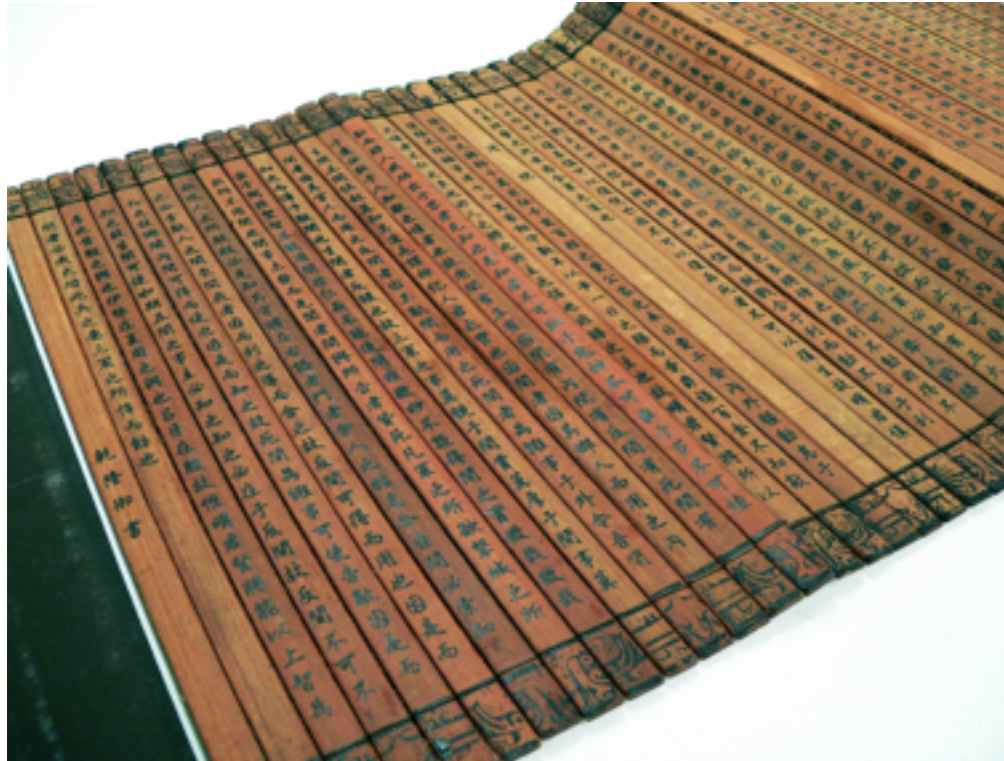


Automating Data Extraction from Chinese Texts

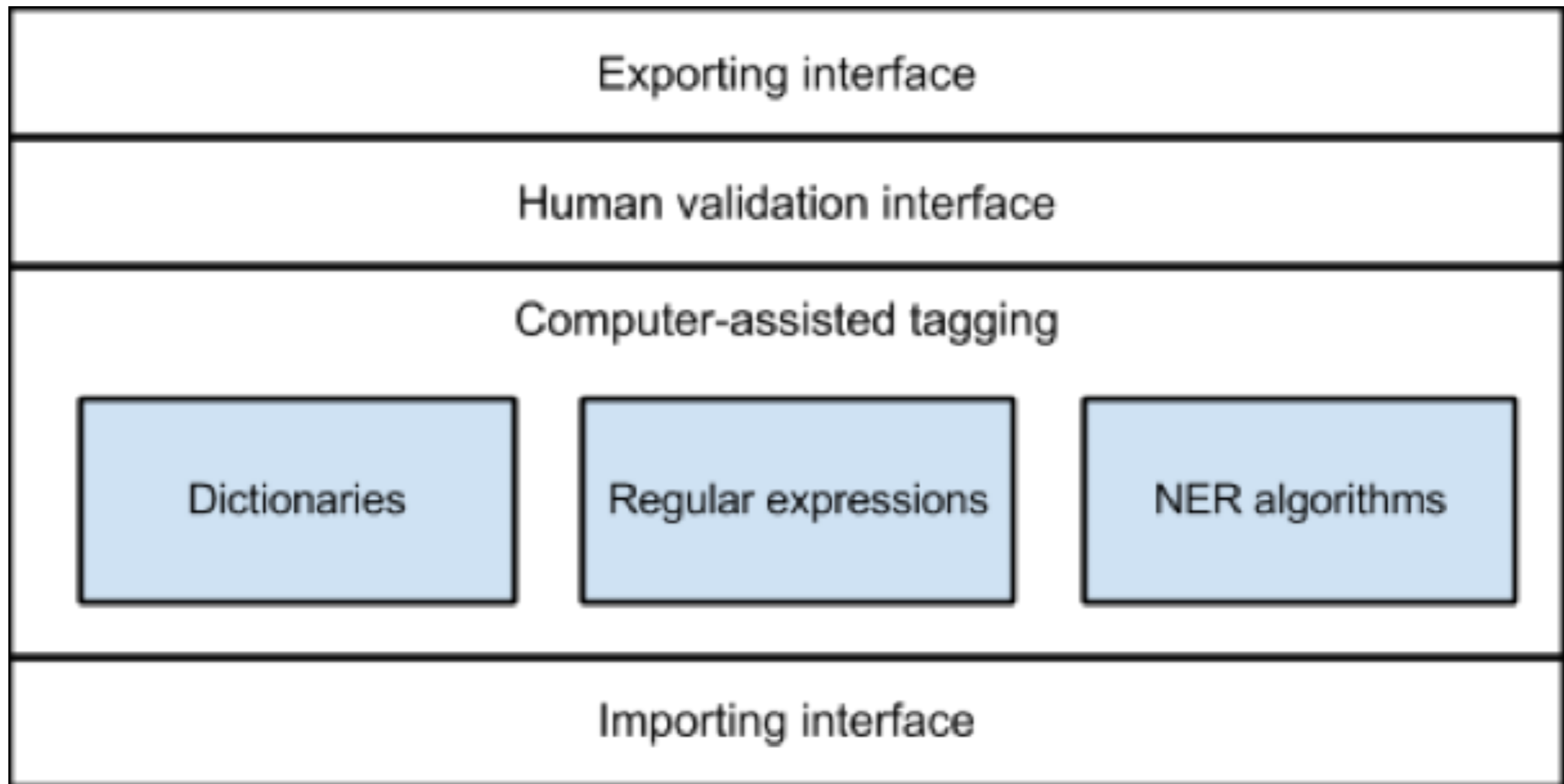
Brent Ho
LIAS, Leiden University



2200 years

By Academic and cultural institutions, commercial vendors, and individuals

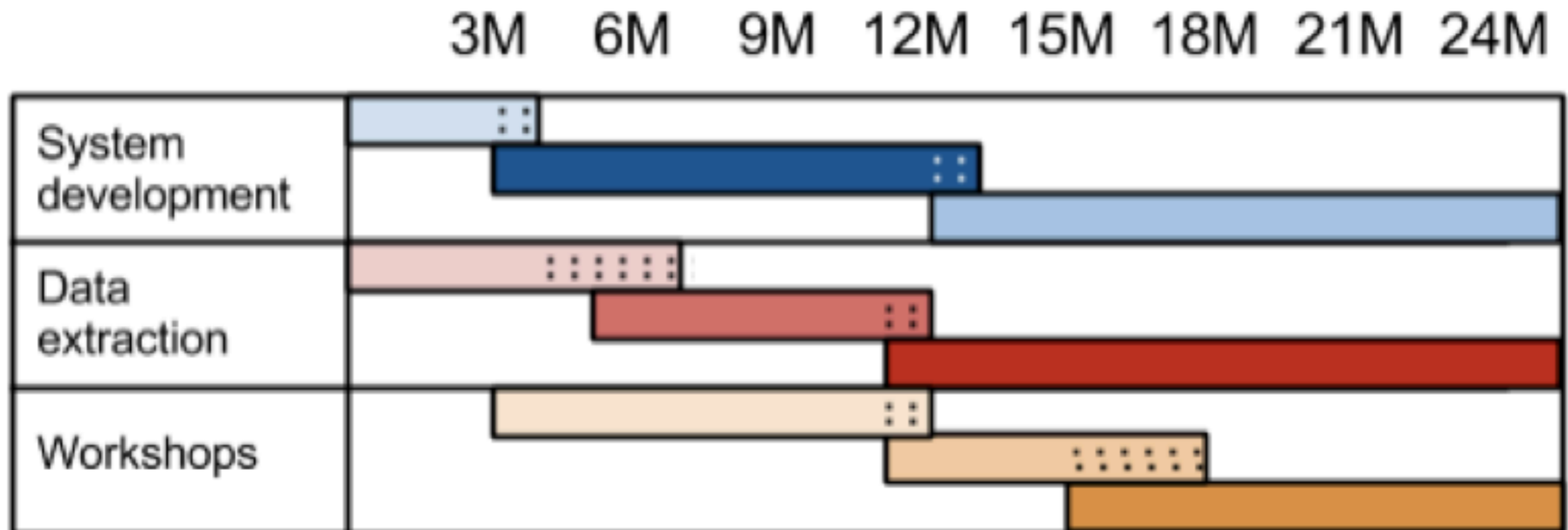
The system components of MARKUS



Team members



Project schedule



Three major outputs

- (1) an open platform for the tagging and extraction of data from Chinese historical documents
- (2) data extracted from local gazetteers
- (3) workshops and supporting documentation.

Q & A