

What We Can and Cannot Do with the Computer:

The Potentials and Limits of Computational
Chinese Studies from a Librarian's Perspective

Jidong Yang
Stanford University

What is digital humanities anyway?

- Nowadays it is very hard to find any humanists or social scientists doing research without using a computer and digital resources. But are they all digital humanists? — Apparently not.
- Most available digital resources for Chinese studies, such as full-text databases, are not very different from print-era textual indexes/concordances in nature and usability. Current scholarship based on those digital resources mostly adopts traditional methods of reading, interpreting, and analyzing.
- The threshold for digital humanities: algorithm (applied to GIS, data mining, network analysis...). The main process of research is not reading, understanding, and keyword searching of texts, instead, it is machine-processed automated computing, sorting, mapping, and analyzing of large quantity of data.

Library and DH; the Stanford Model

- Stanford University Libraries' CIDR (Center for Interdisciplinary Digital Research) and its roles: 7 full-time librarians with strong background in humanities/social sciences and IT. Curation and preservation of data. Daily consultation services for faculty and students. Workshops and training. Most importantly: direct engagement in selected faculty digital projects.
- Stanford University Press (now a part of the university library system) started publishing digital projects, all of which received technical support from CIDR. Thomas Mullaney's *The Chinese Deathscape* is the latest such publication.
- East Asia Library's role in supporting DH: providing raw data (through purchasing, digitization, web archiving, etc.) and making them accessible to digital research projects.

Controversies and obstacles

- Is DH already the mainstream of Chinese studies? Apparently far from it. Of the 89 sinological papers published in the *Harvard Journal of Asiatic Studies* and *T'oung-Pao* during the last five years, only 1 can be considered a DH work.
- Copyright infringement exists in many digital research projects, causing threat to the survival of commercial publishers and companies that are producing best-quality primary research sources (printed and digital) for the field.
- Intolerable high percentage of errors in researcher-contributed raw data, metadata, OCRed full texts, punctuation and annotation of texts, character coding, markup tags, maps, ...
- Successful digital projects are expensive and resource-consuming.
- Sustainability of any digital scholarship projects is always a big challenge. No easy solutions on the horizon yet.

To sum up

DH is cool, but we also need to keep a cool head.