

*Building an Infrastructure for  
Digital Humanities: Issues and  
Considerations*

**Peter Zhou**  
周欣平

University of California, Berkeley  
October 8, 2009

# E-humanities

- E-science/e-humanities: large cyber-infrastructure to facilitate interdisciplinary research and data in a networked environment

Terms: **cyberinfrastructure, e-Infrastructure, e-research**

# Components

- A. Human sphere (people and cross-disciplinary collaboration, networking & partnerships)
- B. Implementation streams (cyberinfrastructure, constructs, discovering tools, implementation platform)
- C. Data (glue of collaborative research) such as data net, documents, publications, composite objects and links

# What is data?

- Data has a wide variety according to disciplines, such as
  - Specimens in biology
  - X-rays in medicine
  - Mass media in social sciences
  - Numbers in mathematics and statistics
  - Artifacts in archaeology
  - Sensoring data in earth sciences
  - Images in anthropology
  - Archival texts in history and literature
- Data is where the library comes in

# Library and Data

- Data selection & linking (Google cannot do hyperlinks; It requires library, text-to-text links, database-to-database links)
- Data sharing (licensing and copyright)
- Data storage (data lab and data center)
- Interoperability of data such as those in many databases
- Create single point access to many databases, even cross language barriers.

# Data value chain

- Legitimization
- Dissemination
- Curation and preservation

# Goals of e-humanities

- Bring network revolution from culture and commerce to research;
- From finding a shoe on the web to finding an archeological object;
- From booking and viewing hotel room to viewing the architecture of a temple;
- From chatting and dating services to scientific networking and online communication for large scale research on humanities



# Library in e-research

- Library will interject itself in e-research and provide infrastructure for a long time for preservation, citation, location, structure and discovery.
- Library glues e-research together and provide the whole picture.
- Library plays a pivotal role in data-centric e-research today.



# Directions in E-science/e-humanities

- Interdisciplinary
- Discovery tools revealing people, data and relationships
- Infrastructure to serve the global community, not just the campus
- Data-intensive

# Initiatives in Berkeley's Starr East Asian Library

- To create an infrastructure to facilitate research and scholarship on East Asia
- To function as a major hub for collecting, storing, and disseminating information digitally on East Asia

# Building Content

- E-books and e-journals are becoming the standard format for publication and research in Chinese studies. Numerical, GIS, and other types of data delivered electronically are critical to research in humanities and social sciences and professional studies, particularly in the fields of economics, finance, trade and banking.
- The Starr Library already owns or has subscribed to more than 700,000 e-books and more than 6,000 full-text e-journals.

# A New Digitization Project



# The Asami Collection and Korean Rare Books

<b>Collection</b>	<b>Titles</b>	<b>Volumes</b>	<b>Pages (est.)</b>
<b>Asami</b>	<b>900</b>	<b>3,400</b>	<b>510,000</b>
<b>Other</b>	<b>1,500</b>	<b>4,500</b>	<b>675,000</b>
<b>Total</b>	<b>2,400</b>	<b>7,900</b>	<b>1,185,000</b>



# Key Components of the Project

- Digitizing all of the rare Korean materials, including the Asami collection, currently held by the Starr Library.
- Providing complete metadata to enable easy and universal access through both the open web and library OPACs.
- Mounting the digitized materials on the Internet in UC Berkeley and Korea University

Text Archive > Test books from California > asamitest

View the book



- [Flip Book](#) (1.3 MB)
- [Flip Book \(beta\)](#) (17 MB)
- [PDF](#) (2.6 MB)
- [B/W PDF](#) (2.6 MB)
- [Full Text](#) (1.5 KB)
- [DjVu](#) (2.1 MB)

All Files: [HTTP](#)

[Help reading texts](#)

asamitest

Possible copyright status: NOT\_IN\_COPYRIGHT  
 Digitizing sponsor: [Internet Archive](#)  
 Book contributor: starr library  
 Collection: [richtest](#)

Description

bound item 1

Reviews

Be the first to write  
Download

Selected metadata

Identifier: asam01testrich  
 Mediatype: texts  
 Noindex: true  
 Scanningcenter: rich



# Interactive and archiving features

- Attachments & comments
- Editorial oversight
- Scholarly annotations and reviews
- Bookmark
- Report errors
- Digital archiving and preservation

Questions?

谢谢!