

# Creating a Free-to-Read International Digital Library – Five Years Later

Erika C. Linke

Carnegie Mellon University  
U.S.A.

# Overview

- Background
- Challenges
- Current Activities
- Final Thoughts/Reflections on the Future

# Background

- Create a test bed for computer science research in
  - Machine translation
  - Automatic summarization
  - Data mining
- Collection content
- 1.6M volumes by September 2007

# Challenges

- Getting enough content from the U.S.
- Copyright

# Challenges continued

- Reluctant to send content of US partners before we had full experience of
  - 1) shipping
  - 2) scanning quality
  - 3) metadata quality

# More Challenges

- Tremendous need for content
  - Container shipments
  - Some content was to be returned; other content was not to be returned
  - Gap filled by Chinese books
    - Change in Chinese copyright law created the opportunity to scan massively. Content only available to students in China

# More Challenges Continued

- Books don't travel well
- Not all scanned content was accessible

# Underestimated...

- Bandwidth - Internet Access
  - Use OCLC for metadata was entirely feasible
  - Too many files to move using the Internet
- Experience/Education of employees hired
  - Ability to read and understand English
  - Ability to assign subject headings

# Underestimated...

- Fundamental importance of accurate metadata
  - Retrieval
  - Opinion leaders
- Search and retrieval
  - Doesn't fit with initial research agenda
  - Finally dawned on researchers that a simply structured search on limited elements did not satisfy

# Underestimated...

- Disciplinary challenges
- Different styles of group work
- Communication

# Opportunities

- Bibliotheca Alexandrina partnership
- *Chandhamamma*
- Continued interest of funding agencies
- Interest from other organizations who define their role internationally

# Current and Future Activities

- Data clean-Up
- Data integration
- Discovery
- Grow collection/new targets

# Final Thoughts/Questions

- Diverse content
- Impact of large-scale digital projects
- Interdisciplinary collaboration
  - Services
  - Development
- Genuine and deep commitment to continue working together