C. Making East Asian digital resources web-scale: infrastructure for East Asian collections

Jim Cheng, UC San Diego Libraries
Major Issues:
Evolving models of the digital library service

   - Standards (Metadata, Unicode, CJK OCR ...... How about the oversize and pictorial materials? )
   - Fully Integrated Access (with other Google Products; Google Book Search API; CJK Unicode/Full Text Search with no Viewers and Plug-ins; Limited CJK Unicode Characters? Different Browsers? )
   - Rights management (Applied to American Copy Right, with CJK materials?)
   - Perpetual access/Preservation (Digital copies for the Involved Institutions)
Major Issues:
Evolving models of the digital library service

2. Consortium Model (UC and NERL/YALE):
   – Purchased/owned
   – Lack of the Consistent and systematic Standards
   – Obsolete software and evolving hardware problems?
   – Sustainable budget and staff support for maintenance and user services?
   – Preservation
Major Issues:
Evolving models of the digital library service

3. Vendor Mirror Site Model (EastView, ChinaMaxx, and Wanfang Data):

– Problem with the mainstream standards (Metadata, SFX/Open URLs, Access authentication……)

– Access to CJK contents (Unicode fonts/Viewers/plug-ins, browsers, operation systems, future MetaSearch issue ……)

– Perpetual Access (Tailored rights to purchased materials when the subscription is cancelled)

– Digital Preservation (A trusted digital preservation depository)
Major Issues:
Evolving models of the digital library service

4. Vendor Direct Service Model (CiNii, KNS, and UDNDATA):
   – Problem with the mainstream standards
   – Access to CJK contents
   – Perpetual Access
   – Digital Preservation
Major Issues:
Evolving models of the digital library service

5. ArtStore Model (for digital images used by multiple institutions):

- Standards for digital images (E.g. Image Resolution: 2,400 dpi; different Metadata standards comparing to MARC 21, no CJK characters in Metadata)
- Fee-based bibliographic and content utility (not a Open Source)
- Advanced display tool
- Required training and download the Off-Line Viewer
Major Issues:
Evolving models of the digital library service

6. Local Hosting Model (LC, UCB, USC, and others):
   - Locally created and managed CJK digital contents
   - Open access without required clients, viewers and Plug-ins
   - Mainly graphic and map materials, but small amount of books and journals, often rare materials.
   - No CJK OCR application and no full-text search
   - Digital Preservation
Recommendations

1. Understand and Apply the mainstream standards with the unique focus on EA Digital resources:
   - **Safe Operations**: Redundant Installations with different network paths; High-speed Machines on networks with large bandwidths; Accessible from Internet2-connected sites 24/7; ......
   - **Information Security**
   - **Access Authentication/Mechanisms**: IP control and accommodate proxy server and VPN, future standards: federated identity management and privacy protection-Shibboleth; discourage Client/Password; overly Logoff for simultaneous users.......
   - **Data Integrity**: Currency, completeness, Retracted/missing materials, corrections .....  
   - **Quality of Digital Images**
Recommendations

- **Data Presentation**: http, https, XML, MPEG, MP3......, OpenURL, IP control, no login/password screen, ......
- **Perpetual Access**: Access to the previously subscribed materials
- **Preservation**: Digital preservation depository with OAIS and TRAC standards, such as DAMs and DPR at UCSD and CDL. Not print, CD/DVD and other formats,
- **Inter-Library Loan**: print and digital copies with limited access and time
- **Statistics**: for searches, sessions, turn-aways, and article downloads
- **Downloading Record Format**: Z39.80
- **Unicode**
- **Metadata**: MARC 21, batch loading/XML file, aggregated with the bibliographic utilities: OCLA WorldCat
Recommendations

2. **Integrate access of East Asian digital resources with other mainstream digital resources** (1. configure the public terminals, integrate with other resources in the E-resource lists, 2. bypass the ID/Password by using INI, PHP, and Registry programs, 3. MetaSearch: Federation/A single query—Z39.50, SRU/SRW, and NISO standards; Aggregation/A single index—harvesting metadata or crawling the full text contents, ...... )

3. **Governing Law as the contract term**

4. **Sustainable model for hosting EA digital resources**

5. **Cooperation with the vendors for the local digitization projects related to the CJK textual materials**