Aiming at Quality and Coverage Combined:
Blending Physical and Virtual Union Catalogues

Janifer Gatenby, OCLC PICA

The European Library (TEL) 29-30 April 2002

Die Deutsche Bibliothek, Frankfurt Am Main
Straight for the Jugular

- Libraries are wasting their time now that we have Internet, the Web and search engines like Google
Web Search Engines are great except for:

- **Precision** –
  - Retrieves too much: Full text indexing: Relies on ranking.
  - Retrieves irrelevancies
  - Indexing not precise because lack of metadata; e.g. no controlled subject searching

- **Quality** –
  - Retrieves poor quality materials; ranking only partly compensates

- **Recall** –
  - Relevant material often buried
  - Broken links
  - Web services and contents of databases
  - **What about the resources not on the Web?**
Belief in Metadata

- Cataloguing / Metadata
  - Selection of quality materials for treatment
  - Precise indexing = better retrieval
    - Descriptive cataloguing
    - Subject cataloguing
    - Authority control
- Authoritative
  - Description – positive identification;
  - Declaration of Identification not dependent on address

- So much information to cover; so much to do

Cooperation
Metadata Retrieval Models

- Distributed
  - Protocol based, Z39.50 / ZING/SRW / XML Query
  - No centralisation
- Distributed data; centralised index
  - Protocol – OAI harvest
- Centralised
  - Union Catalogue
Metadata Retrieval - Distributed

- Z39.50 model
  - Distributed virtual catalogue
  - Common searches
  - Combined results
Distributed Model

UI  Z39.50 client
Z39.50 Search > a search engine

- Precise searching
  - e.g. title only
  - Boolean operations, nested etc.

- Results set handling
  - segmentation
  - sort & duplicate removal
  - refine search, no cookies

- Browse
Z30.50 Search > a search engine cont.

- **Retrieval options**
  - brief or full records
  - can specify content

- **Choice of format**

- **Extended services**
  - SDI - saved query and schedule
  - Saved results
  - Item order, and update

- **Linked – electronic & physical**
Z39.50 Advantages

- Multi-target searching
  - One user interface; one search
  - different platforms (UNIX, NT, IBM etc.)
  - different database systems (relational, network)
  - different database models
- Searching based on abstract concepts
  - Title, i.e. not database columns
- Can combine results from diverse databases
  - Common record syntax (usually MARC)
- Reuse of results
- Follow on delivery – electronic & physical
Z39.50 Drawbacks

- Complex and difficult to implement
- Too many options
  - Interoperability problems without profiles
- Interoperates with web but........
  - Not fully web compatible
- Has not achieved widespread acceptance
- Contenders
  - ZING SRW / SRU
  - XML Query / SQL like
  - OAI
  - OpenURL
Better WWW Integration – ZING

- SRU = Simple URL…….. in web browser
  - Target can include style sheet indicating display
  - Very thin client
- SRW = more complex
  - SOAP wrapper, remote procedure call
  - Send and receive XML documents
- Common elements
  - remove statefulness; no init
  - Only 1 record syntax = XML
  - text passed not encoded data
  - CQL based on CCL search
  - Inherits search modelling of Z39.50 & Bath
Metadata Retrieval - Centralised / Distributed

- OAI Model
  - Data distributed
  - Harvested
  - Centralised index

- Google – like
- More efficient than distributed model
OAI Model

UI

Index

OAI client
OAI Model - Assessment

- ADVANTAGES
  - Common index - power
  - Common index - speed
  - Simple model for update

- DRAWBACKS
  - Designed for documents not metadata
  - Metadata can change when updated, e.g. authority control
  - Additions; no maintenance; no deletion
  - No allowance for alignment; diagnostics & IDs
Metadata Retrieval - Union Catalogues

- Centralised metadata index
  - Shared cataloguing
  - Discovery – rich searching
- Locations – loan and ILL
- Links to text; abstracts; reviews; etc.
- Designed for metadata
- Complete update model
Union Catalogue Model
<table>
<thead>
<tr>
<th>REAL</th>
<th>VIRTUAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Many accesses</td>
<td>Lowest common den.</td>
</tr>
<tr>
<td>Speedy – bespoke srch</td>
<td>Slower</td>
</tr>
<tr>
<td>Precision</td>
<td>Only small sets</td>
</tr>
<tr>
<td>- duplicate detection</td>
<td>Dep. Profile adherence</td>
</tr>
<tr>
<td>- Consistent indexing</td>
<td></td>
</tr>
<tr>
<td>Recall</td>
<td></td>
</tr>
<tr>
<td>- Authority control</td>
<td></td>
</tr>
<tr>
<td>Delivery</td>
<td></td>
</tr>
<tr>
<td>- ILL; text links</td>
<td></td>
</tr>
<tr>
<td>Multiple views</td>
<td></td>
</tr>
<tr>
<td>Resource</td>
<td></td>
</tr>
<tr>
<td>- Data mining</td>
<td>No maintenance of centralised index</td>
</tr>
<tr>
<td>- Collection development</td>
<td></td>
</tr>
<tr>
<td>- Authentic source</td>
<td></td>
</tr>
</tbody>
</table>
Pan Union Catalogues

- Maximize the size of union catalogues
  - Large physical union catalogues where possible
  - Protocol links where not practical
  - Links with other large union catalogues
Pan European Model

UI

Index

Document Supplier

Z39.50 UC

Regional UC

Regional UC

National UC

National UC

ILL Services

Local Cat

Local Cat

Local Cat

Local Cat
EUCat

- Centralised index
- ILL from regional and national Union Catalogues
- Local view of EUCat
  - Tailored
- Clustering of ‘duplicates’
  - Different language versions, different subject headings and classification
  - Presentation dependent on affiliation of user, example German record, German holdings displaying before other holdings
- Clustering of authorities
EUCat

- 3 levels of partnership
  - Full – contribute metadata and holdings
  - Contribute metadata but no holdings
  - Remote – allow Z39.50 access

- Currently:
  - GBV
  - NCC
  - ZDB

- Hopefully Soon
  - EU holdings from WorldCat
  - Other negotiations in progress
Extended WorldCat (XWC)

- OCLC WorldCat currently 49 million +
- EUCat currently 30 million +
- Other nodes
- Bibliographic and Authority Data, Library Directory – Pooled Commodity
- Holdings and Services
  - Local to the Nodes
  - Cross node services and holdings discovery
- Statistics and Authentication
  - to support multiple financial arrangements
Global Resource…Regional Autonomy

**OCLC WorldCat**
- Replication: Bib, Authority and Library Directory Data
- Search: Holdings, Bib, Authority & Content
- Clearing House

- **Dublin Node**
  - + Holdings
  - Replication
  - Search
- **European Node**
  - + Holdings
  - Replication
  - Search
- **Other Node**
  - + Holdings
  - Replication
  - Search
- **Local System**
  - Replication
  - Search

**Services**
- Z39.50

**Union Catalog**
- Z39.50
- Replication
- Search
- Local System
Which standards - Searching?

- Distributed searching
  - Z39.50
  - ZING SRU
  - Bath profile
    - Bibliographic
    - Holdings
    - Authorities
    - Cross domain

- FRBR potential
  - Better presentation and navigation
Which standards – Update?

- Update – Seamless population of union catalogues
  - OAI
  - Extended with UCP?
    - Diagnostics
    - Record structures for maintenance and deletion

- Cataloguing
  - FRBR potential – copy cataloguing at levels; better management of rights
  - DC, Mods, ONIX – simplified cataloguing
Which other standards?

- Linking – OpenURL
  - Full text
  - Related materials, e.g. reviews, citation indexes
  - Related services, e.g. order
- ILL – ISO ILL
- Circulation – NCIP
- Directories – ISO 2146
  - ILL sources
  - Reference queries
  - Collection description
- Reference queries – NISO working committee CDRS
  - Protocol for exchange of queries
  - Data schema
Grazie

Vielen dank

Merci

Muchas gracias

Dank u wel

Thank you

Thank you

G'DAY