Strathprints Institutional Repository

Nicolaides, Fraser and Macgregor, George (2004) Interoperability : the performance of institutional catalogues & strategies for improvement. In: Hyper Clumps, Mini Clumps and National Catalogues: Resource Discovery for the 21st Century, 2004-11-11, British Library Conference Centre. ,

This version is available at http://strathprints.strath.ac.uk/57714/

Interoperability: the performance of institutional catalogues & strategies for improvement

Fraser Nicolaides,
M25 Systems Team

George Macgregor,
Centre for Digital Library Research,
University of Strathclyde

interop

JISC

The interoperation of catalogues in large-scale, distributed search environments

The study: a comparison of the performance of physical and virtual union catalogues

Non-system architecture issues (semantic interoperability):

- System-level variations in the ability to process search queries
- The impact of disparate cataloguing and indexing practices on recall and precision

Recommendations and other remedial activities

interop

JISC

Scope of the study: 1

For the test methodology, search results and associated discussion, see:

'A comparative study of the performance of COPAC and selected independent Z39.50 servers', Issue 1 (June 2003); available at:

http://ccinterop.cdlr.strath.ac.uk/documents/WPA_server_tests_issue1.pdf

interop

JISC

Scope of the study: 2

Virtual union catalogue: 6+ CURL libraries

- All contribute records to COPAC

Physical union catalogue: COPAC

- Subsets of the database were searched (*i.e.* restricted to above libraries)

All of the catalogues were accessed using the Z39.50 information-retrieval protocol

interop

JISC

Variations in performance

Consistent differences in performance were identified. These were attributed to:

- The currency of the physical union database;

- The availability of records that describe (and may enable access to) electronic resources;

- The definition and processing of search types ('author', 'title', *etc*.); and

- The cataloguing and indexing of bibliographical entities.

The 'administrative' issues

Currency of the physical union database

- The time delay in submitting records to COPAC

Access to electronic resources

- The non-submission to COPAC of records describing (licensed?) electronic resources
- An administrative policy not extended to the institutional Z39.50 server

interop

JISC

The definition of search types: Bib-1

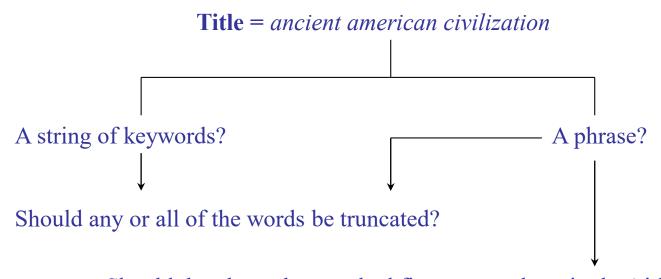The considerable variations between *all* of the systems in their support for Bib-1 attributes

The use of Bib-1 attributes to define precisely the abstract concepts of 'author', 'title', 'subject', *etc*.

- 'Attribute set Bib-1 (Z39.50-1995) semantics'; available at:

  ftp://ftp.loc.gov/pub/z3950/defs/bib1.txt

- 'Bib-1 attributes set'; available at:

  http://www.loc.gov/z3950/agency/defns/bib1.html

interop

JISC

What do we mean by a 'title' search?

Title = *ancient american civilization*

A string of keywords?   A phrase?

Should any or all of the words be truncated?

Should the phrase be matched first or anywhere in the 'title' field(s)?

All of these (and other permutations) are legitimate forms of 'title' search.

Their potentially highly variable impact on recall & precision.

interop

JISC

Example definitions of the 'title' search type

The use of Bib-1 attribute combinations to give formal semantic definition to any search type. For example:

Attribute type	Title search - keyword		Title search – exact match	
	Value	Attribute	Value	Attribute
Use (1)	4	Title	4	Title
Relation (2)	3	Equal	3	Equal
Position (3)	3	Any-position-in-field	1	First-in-field
Structure (4)	2	Word	1	Phrase
Truncation (5)	100	Do not truncate	100	Do not truncate
Completeness (6)	1	Incomplete subfield	3	Complete field

interop

JISC

Selective system support for Bib-1 attributes

Multiple attributes – multiple attribute combinations

The difficulty of finding attribute combinations shared by all of the tested systems

Default behaviour – the server-end (catalogue) replacement of unsupported attributes

The definition of search types: a solution

Z39.50-server vendors should support a common suite of attribute combinations (basic search types)

The availability of formally published 'profiles'

The core profile for bibliographic applications:

'The Bath Profile: an international Z39.50 specification for library applications and resource discovery', Release 2.0 (February 2004); available at:

http://www.collectionscanada.ca/bath/tp-bath2-e.htm

interop

JISC

Implementation of the Bath Profile

Vendor reluctance…

- A recognition of the potential communal value of compliance;
- The limited scope of their business objectives;
- The cost of rebuilding indexes;
- The perceived instability of the Profile.

…and customer ignorance

- Of profiles;
- Of the Bath Profile;
- Of the Areas and Levels of the Bath Profile relevant to them.

interop

JISC

Cataloguing and indexing the bibliographic entity

Two sub-issues:

- The bibliographical completeness of the catalogue records;
- The indexing policies that effectively determine how those records are accessed.

interop

JISC

The bibliographical record

A distinct advantage of physical union catalogues:

- For any bibliographic entity, there is the **potential** to derive index entries from records submitted by multiple contributing institutions.

The cumulatively enriched COPAC records. In particular, the notable presence of comprehensive added entries (author; title; series…) and subject headings.

interop

JISC

Indexing policies

The policies that determine:

- Which (sub)fields are indexed; and
- To which access points they are mapped (which (sub)fields are used to create the 'title' index(es)?).

The impact of index functionality on the potential to support Bib-1 attributes. For example:

- If the 'title' index does not support first-in-field matching, then the Z39.50-server cannot process queries that specify the corresponding Bib-1 Position attribute.

interop

JISC

Recommendations: 1

Any library that wishes to participate effectively in virtual union associations must detail to their system vendor the measure of semantic interoperability they require from their Z39.50 server.

- Particular reference may be made to the constituent parts of the Bath Profile.

COPAC has already attained compliance with Release 1.1 of the Bath Profile.

interop

JISC

Recommendations: 2

A conservative view:

> "Variations in cataloguing and indexing policies [are] the product of historical and local requirements and contingencies, the legitimacy of which [..] should not be challenged. It is conceivable that an increasing awareness of the impact of local policies on the performance of distributed systems will generate concerted remedial responses."

> F. Nicolaides 'A comparative study…' (2003), p.41.

An alternative view…

interop

JISC

The question:

Improving Interoperability in Distributed & Physical Union Catalogues through Co-ordination of Cataloguing & Indexing Policies

Report available:
http://ccinterop.cdlr.strath.ac.uk/documents/CCICatInterop.pdf

Scope & Background (cont.)

CAIRNS project: Z39.50 based Clump

- eLib Phase 3
- Originally comprised SCURL catalogues, now the Cooperative Information Retrieval Network for Scotland

Not just about building services for users…

- Focus on interoperability
- Technical barriers (conformance, availability, etc.)
- Cataloguing barriers

Formation of the CAIRNS Catalogue Issues Working Group:
liaising with Cataloguing & Indexing Group in Scotland (CIGS), SCURL

interop

JISC

Scope & Background (cont.)

CAIRNS & CAIRNS Catalogue Issues Working Group

- Set of mappings to standard CAIRNS clump indexes (USMARC & UKMARC) produced by CAIRNS
- Each CAIRNS library produced mappings for the indexes used by their Z39.50 server
- Enabled identification of convergent and divergent practices
- Common standard for cataloguing & indexing in Scotland = enhanced interoperability of metadata content and supported wider aims of SCURL

Mechanical & procedural changes were identified:
Short Term vs. Long Term

Benefits reaped…

interop

JISC

CC-interop: B3

CC-interop: Deliverable B3 sought to build on CAIRNS work

• Much has changed since CAIRNS: MARC21, ISBN, Bath Profile conformance, global access to local records, etc.

Interoperability: chasing the rainbow?

• B3 aim to agree a similar set of guidelines and appropriate strategies for improving interoperability and to reflect recent LIS developments

Superior method: Fraser's work, changes at University of Hull

CC-interop: B3 Methodology

2 one day workshops: London & Glasgow
- **London:** invitations sent to CURL and InforM25 cataloguers
- **Glasgow:** invitations sent to SCURL and Scottish FE sector

- Revised/generic CAIRNS guidelines distributed: home work
- Presentations used to stimulate open discussion

Discussion tape recorded

⬇

Transcribed to produce a report

⬇

Distributed to attendees for comment and correction

50+ in attendance from UK cataloguing and systems fraternity!

"Hyper Clumps, Mini Clumps and National Catalogues: resource discovery for the 21st century",
11th November 2004, British Library, London

JISC

Recommendations

General outcomes:

• **Consensus:** prescriptive guidelines essential to assist improvement of interoperability, raising awareness among senior management

• **People interoperability**

• CAIRNS-style guidelines inappropriate (though not without local value)

   Globalisation of cataloguing demands guidelines for a wider, active and nationally co-ordinated approach to improving interoperability

Recommendations and strategies under 4 headings:
   • Collaboration within distributed or physical union catalogues
   • Standards
   • Strategic Developments
   • End Users

JISC

Collaboration within distributed or physical union catalogues

1. Consortia of libraries contributing to union catalogues should develop prescriptive guidelines
2. Catalogue scope, content, accounting for local and global needs
3. One level of aggregation is outmoded: *global in one environment is local in another!*
4. Guidelines need to be developed at national and international levels (Anglophone community?). *Identify mechanisms: Full Disclosure? Etc.*
5. Consortia should consider participation in international activities such as the Program for Collaborative Cataloguing (PCC): *reconcile clashes between local/global name and subject headings, etc.*
6. Shared cataloguing service for digital resources: *digital environment = greater choice, thus facilitating greater cataloguing variations*
7. Communication among cataloguers essential (esp. Consortia): email insufficient

interop

JISC

Standards

1. Further development of Bath Profile should encompass scope and content of specified indexes: *enable standard mappings from MARC21 to Z39.50 indexes*

2. Producing guidelines on required Bath conformance: *reducing choice in the consortia would benefit interoperability*

3. Standard rules for index content normalization should be adopted at 'as wide a level as possible': *name punctuation, titles, subjects, inversion of personal names, etc.*
   Use of NACO?

Strategic Developments & End Users

- Consortia and individual libraries should monitor the implementation of FRBR to plan for large-scale machine processing of catalogue data to improve interoperability

  Meeting FRBR model dictates dis-aggregation of existing catalogue record structure and re-aggregation into different structure: *precision needed*

  FRBR implementation: *The pervasive economic argument…*

- Disclose local practices which may affect interoperability for an end-user: *divided opinion…*
- Consortia should agree standard information about each catalogue which should be disclosed as part of union catalogue service: *additional information disclosed at discretion of local catalogue library*

interop

JISC

Further Details

'Improving interoperability in distributed and physical union catalogues through co-ordination of cataloguing and indexing practice'

Available:
http://ccinterop.cdlr.strath.ac.uk/documents/CCICatInterop.pdf

'A Comparative Study of the Performance of COPAC and Selected Independent Z39.50 Servers'

Available:
http://ccinterop.cdlr.strath.ac.uk/documents/WPA_server_tests_issue1.pdf

Contact: george.macgregor@strath.ac.uk
F.Nicolaides@greenwich.ac.uk

JISC