OCLC Pica CBS
The Central Library system

A generic solution for:
- creation and maintenance of union catalogues
- controlled document ordering and delivery

Fourth generation → CBS4

Technical data

- Hardware: SUN (preferred)
- OS: UNIX (Solaris)
- RDBMS: Sybase
- Language: C – C++
- Bibl. database: Generic Pica+ format (Sybase as record manager)
- Index: UNIX file with real time update extension
- Other data: SQL tables
Main functions - components

- Search – retrieve – index: PSI
- Cataloguing: CAT
- InterLibrary Loan: ILL
- Batch import, incl. duplicate detection and record merging: IMP
- Batch export: BibServ
- Z39.50 target: TOLK
- Z39.50 origin: iPort
CBS4 architecture

Features

- Stateless servers
- Limited number of processes
- C - C++
- Posix compliant
- Multilingual:
  - four language tables
  - dynamic language switch
Database structures

- Full Sybase implementation for
  - users
  - libraries
  - statistics
  - ILL requests

- Sybase record manager for
  - bibliographic records
  - local records copy records
  - UNIX file for indexes

Current CBS implementations

- Pica CBS: GGC / NCC:
  - 12M records, 20M copies; 800 sim. users
- GBV: GVK:
  - 21M records, 42M copies; 1200 sim. users
- France: Sudoc:
  - 7M records, 17M copies; 800 sim. users
- Hebis: Hebis:
  - 11M records, c. 16M copies; 300 sim. users
- DDB: Iltis:
  - 8M records, c.14M copies; 400 sim. users
Bibliographic formats

- Internal format: Pica+
- Cataloguing format: can be defined – table driven
  - Holland and Germany: Pica3
  - France: Unimarc
  - Asap: Marc21

Bibliographic features

- Integrated database for all records:
  - bibliographic records of all material types
  - authority records
  - ‘private’ records
- Generic format
- Linked record structures:
  - bibl. - bibl.
  - bibl. - authority
  - authority - authority
- Three levels:
  - 0 = shared
  - 1 = local
  - 2 = copy
PSI – The Pica Search and Index engine

- Index engine
- Search engine
- Set admin
- Presentations for patrons
- SRU compliant – DC, XML-Marc

Use of PSI:
- CBS: for indexing and professional searches
- Patron access to Union catalogues
- Creates link to local system OPAC

PSI – Index engine:
- Inverted index (dictionary)
- Inverted links index (bibl. - authority links)
- Flat links index (hierarchical links)
- Real-time update of indexes

Flexible index definition:
- Table driven:
  - Which tags and subfields must be indexed
  - Which index routine must be used
- Multiple indexing:
  - Same tags and subfields with different index routines
- Conditions:
  - Per (logical) file, material type, language
PSI – Search engine

- Dynamic language switch
- Unlimited set size
- Boolean operators
- Proximity, adjacency
- Set combination
- Multiple set sorting (relevance, year, alphabetical)
- ADI

Professionals: command driven
Patrons: menu driven

PSI – Presentations for patrons:

Features:
- Table driven full and short presentations
- Multilingual
- Flexible interface
- XML – SRU interface for local clients
- OpenURL-based links with ILL and circulation modules
CBS – professional presentations:

Features:
- Table driven full and short presentations
- Current presentation tables:
  - cataloguing format
  - Marc21
  - Unimarc
  - labelled
  - ISBD

CAT – Cataloguing

- Windows client: WinIBW – Unix server
- Cataloguing format can be defined:
  - Holland, Germany: Pica3
  - France: Unimarc
  - Asap: Marc21
- Logging for online and offline export:
  - “own”/“all” updates
  - linked records
  - Supported formats: Marc21, Unimarc, MAB, DC(X), Etc.
CAT – Cataloguing functionality

Server:
- user privileges
- record validation:
  - syntactic
  - semantic
  - table driven
- insert/update/delete of records in real time
- insert/update/delete of indexes in real time

Cataloguing client WinIBW:
- expert mode: full screen editing
- templates for coded data
- novice mode templates
- VB scripts for specific operations (linking etc.)

ILL – InterLibrary Loan

Functions:
- Ordering
- Transmission
- Delivery
- Invoicing

Interfaces:
- Librarians: WinIBW
- Patrons: browser: PSI – patron interface
- External clients: OpenURL
- Suppliers: ISO-ILL
Document ordering:

- Different templates for requesting:
  - Non-returnable items: copies of journal articles, book chapters
  - Returnable items: books
- Automatic provider selection:
  - (journal) holding analyses
  - library parameters (daily threshold, type of library, region, preference lists)
- Präsentation of candidate list / rota
  - (optional) rota editing
- Interfaces for patrons and librarians

Transmission functions

The ILL gateway:

- ISO ILL (SMTP / Edifact based):
  - sending request to first / next provider on candidate list
  - support of ANSWER, SHIPPED, CANCEL, FORWARD notification, RECEIVED, RETURNED and CHECKED-IN
- ARTELL format
- SUBITO format
- Link with LBS Circulation module (OUS)
Delivery functions

- Natural language based commands for:
  - browsing, selecting
  - printing slips and claims
  - downloading
  - online reply (SHIPPED, ANSWER)
  - automatic reply via Pica LBS, Abes PEB

- Example:
  - show ‘my’ new requests

- WinIBW interface for librarians

Invoice functions

- Deposit accounts for patrons and institutes

- Charging tables
  - for requesting libraries (local tariffs)
  - for providing libraries
  - per shipment type, preference list, region, local etc.

- Clearinghouse function for requesting and providing libraries
IMP – batch processing of bibliographic and authority records

Functions:
- Record conversion
- Duplicate detection
- Record linking
- Database conversion

Record conversion

- FCV – library:
  - Conversion rules for online and offline conversions
  - Pica+ as the internal format

- Conversions:
  - Syntax-driven definition of conversion rules
  - No programming skills required
**Duplicate detection**

- Selection of possible duplicates: PSI – Search engine
- Evaluation of possible duplicates: Evaluation server
  - Tri-gram analyses
  - Dedicated evaluation rules (1st, 2nd symposium on …)
  - Calculation of similarity values
  - Weigh factors per tag
  - Overall similarity value:
    - 0.90 → duplicate
    - 0.60 → new record
    - 0.60 – 0.90 possible duplicate
- Merging of duplicates:
  - Merging tables on tag level
    - Add / drop / merge tags
    - Define “best” content

**Record linking**

- Batch software for:
  - Linking authority records with bib.-records
  - Generating authority records

**Database conversion**

- Framework:
  - PSI for searching and indexing
  - CAT for validation and record update
  - Site and situation dependent conversion rules
BibServ – Batch export

- Selection
- Sorting
- Layout definition

Selection

- Log-file selection for cataloguing production:
  - batch: ftp – automated process: GTD – Get Title Data
  - Online: Online Update Fetch – OUF
  - Only “own” updates vs all updates on records “in use”

- Selection on PSI indexes
- Dedicated selection index, defined per library
- ID-number – ranges
- Free text selection after primary selection
Sorting

- Multiple levels:
  - Subject groups – alphabetical
  - Author / title – year – imprint
  - Numerical / decimal sorting for classification schemes
  - Free definition of tags and sub-fields to be used for sorting

Layout

- Cataloguing production: in Marc21, Pica, Unimarc etc.
- Accession lists etc.: XML

Innovative use of CBS:

- Union catalogue for journals / periodicals:
  - ZDB (Deutsche Bibliothek), CCNPS (France), CCP (Holland)
- National authority files:
  - SWD, GKD, PND (Deutsche Bibliothek)
- National Bibliography:
  - Dutch national bibliography
- International databases:
  - EROMM (Göttingen)
- Private files:
  - Chinabase; Japanbase (University of Leiden), STCN
Development in 2004:

- **General:**
  - Unicode – UTF8: May 2004

- **Cataloguing:**
  - Marc21 cataloguing: June 2004
  - Meta-data cataloguing: Oct. 2004
  - Cataloguing non-roman scripts: Oct. 2004

- **ILL:**
  - ISO ILL transformer: conversion BER ↔ Edifact: April 2004
  - Tariff differentiation: July 2004