

Inter-Institutional Database Unification

using

Z39.50 and Dublin Core

~ Resource Sharing System for Humanity Researches ~

2003.11.09

@PNC 2003 Bangkok

Shoichiro Hara

(National Institute of Japanese Literature)

1-16-10, Yutaka-cho, Shinagawa-ku, Tokyo 142-8585, Japan

hara@nijl.ac.jp

(<http://world.nijl.ac.jp/~kiban-s/>)

Scenario

I am a graduate student of Osaka City University. My major is “Literature and Gender.” First of all, I tried to find as many materials as possible from the Web.

When I accessed the NIJL Home Page, I found that **not only catalogue database but also other databases such as full texts, images, archives** have related data, and found that **other institutes also have materials**. But I do not well which institute has what databases, then **I have retrieved databases one by one. And each database system has its own operation**. It was time-consuming jobs!

Are there tools that can access databases of several institutes simultaneously and can retrieve information by one operation ?

Objects

- **Each Institute** creates information retrieval systems that hide differences of
 - Database Locations
 - Database Structures
 - Retrieval Methods
- **But do not change existent database systems**
- **Every institutes should be equal partner**
- **As the result**
 - **Users** see databases on the Web as one database

What are the Problems ?

Most Databases are Heterogeneous...

- **Non-portability**
 - Hardware Dependent, Software Dependent
- **Similar but Different Databases**
 - Historical Background, Different Purposes
- **Incompatibility**
 - Different Operations, Non-Interoperability

Similar but Different Databases

DB1

Name	Sex	Address	Birth	Weight	Height	BP	T.Chol
Shoichiro Hara	M	1-16-10, Yutakacho, Shinagawaku, Tokyo, 142-8585, Japan	1957/5/20	170	65	140/85	220

DB2

First Name	Last Name	Sex	DateOfBirth	Street	City
Shoichiro	Hara	Male	20-May-57	1-16-10, Yutakacho	Shinagawaku

Prefecture	Nation	ZIP	Hint	Weight	Chol
Tokyo	Japan	142-8585	170	65	220

Incompatibility

EXAMPLE:

Retrieve records containing term “TOSANIKKI”

By Hitachi Mainframe Computer

1 / B :TOSANIKKI

By SQL Database

SELECT author FROM books WHERE title LIKE ‘TOSANIKKI’

By PAT

region “title” including “TOSANIKKI”

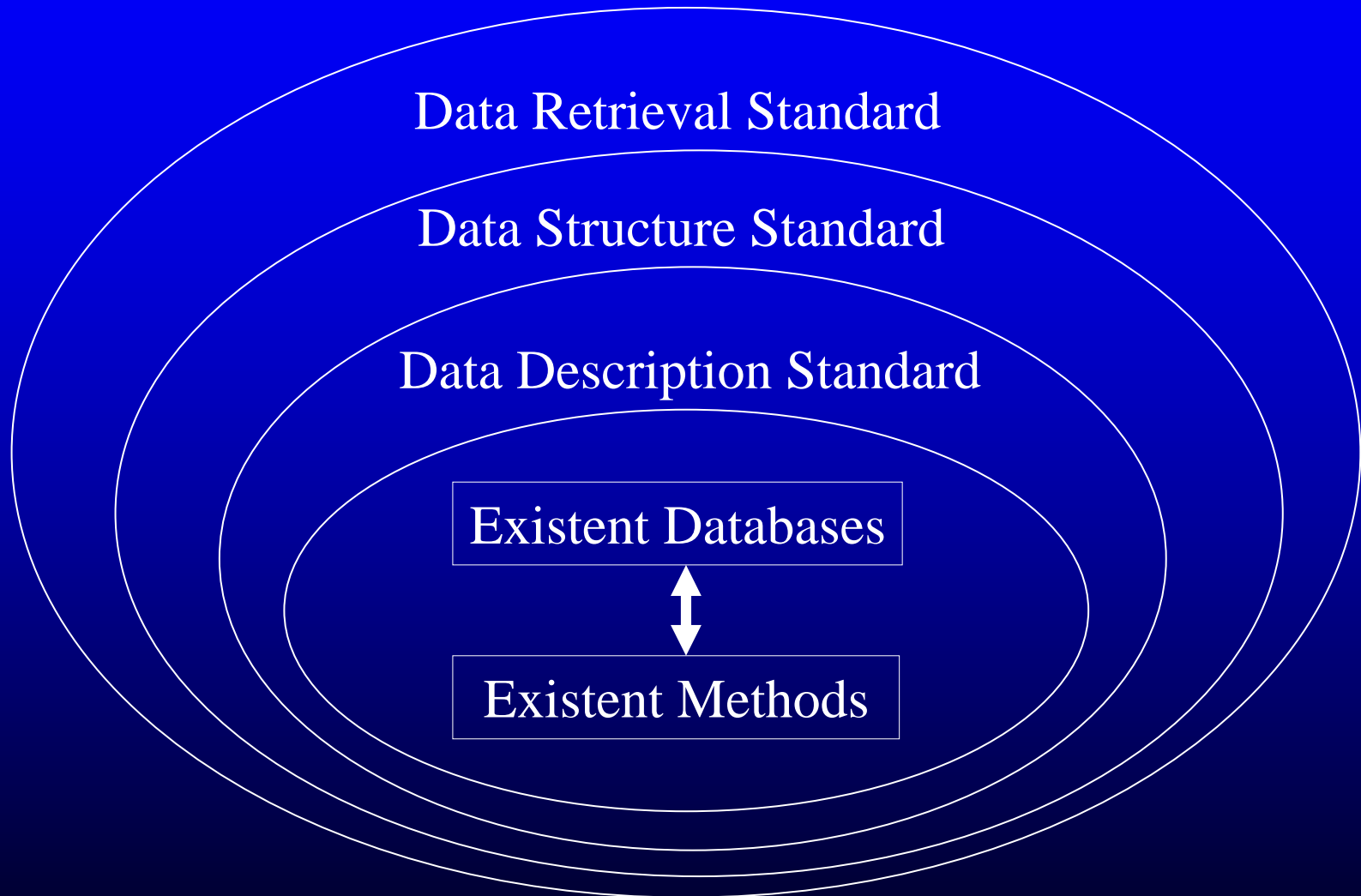
region “book” including %

region “author” within %

Solutions

- **Introducing Standards for**
 - Data Description for Portability
 - Mutual Data Structure for Different Structures
 - Standard Data Retrieval for Compatibility
- **Standardization NOT by**
 - Compulsion
 - Authority

Schema of 3 Steps Standardization



3 Steps Standardizations

1. First Step : Data Description

- **SGML/XML**

2. Second Step: Mutual Data Structure

- **Metadata (Dublin Core, EDI, EAD, TEI etc)**

3. Third Step: Data Retrieval Compatibility

- **Protocol (Z39.50 etc)**



HTML



PDF Text

Step 1

Data Description by SGML/XML

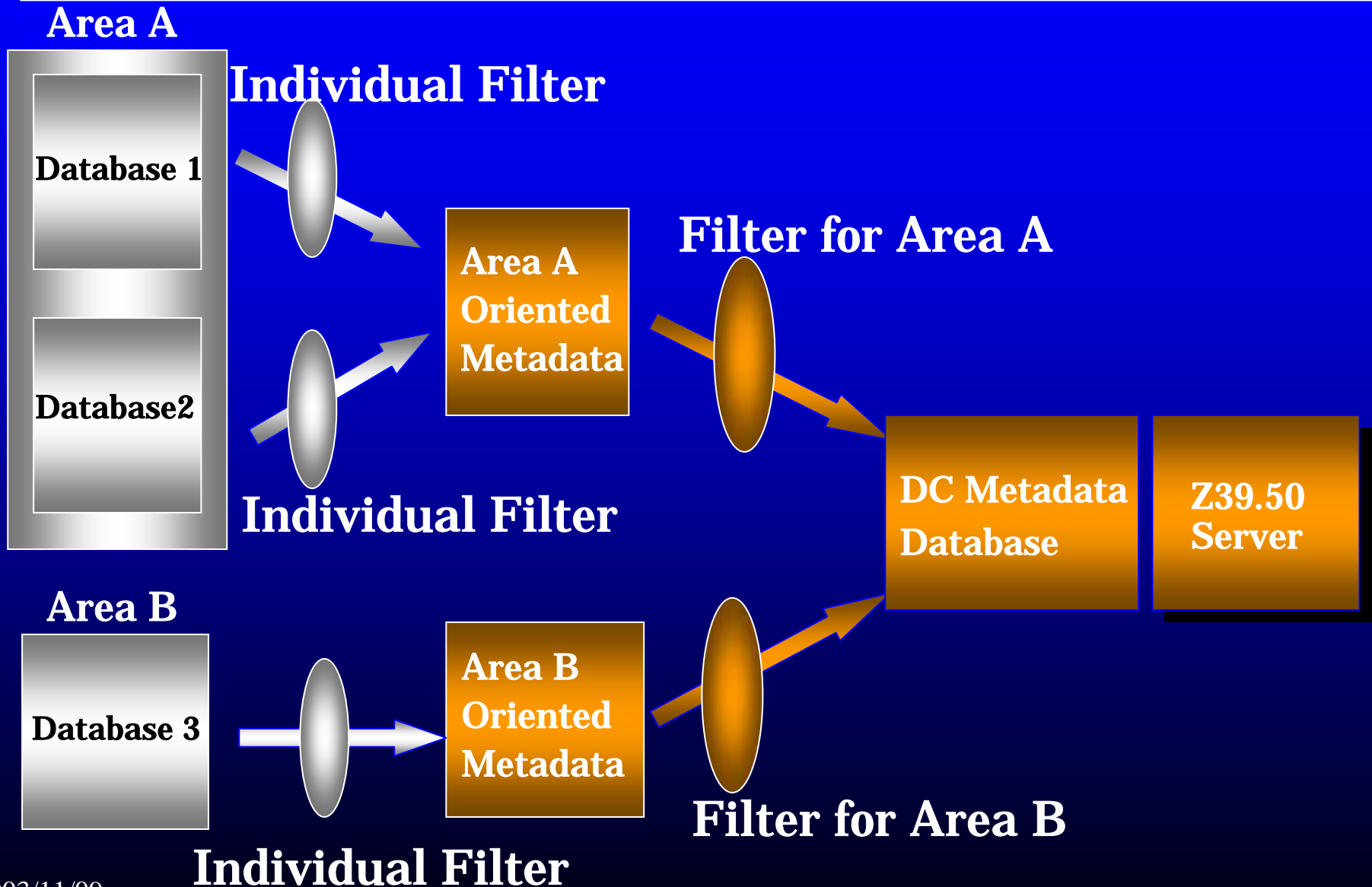
- `<?xml version="1.0" encoding="Shift_JIS" ?>`
- `<!DOCTYPE classics SYSTEM "../dtd/xml-n-classics2.dtd"`
- `>`
- `<?xml:stylesheet type="text/xsl" href="../xslt/classic/honmon_limit.xsl" ?>`
- `<classics>`
- `<text>`
- `<body>`
- `<div0>`
- `<div0.title>源氏物語</div0.title>`
- `<div1>`
- `<div1.title><ruby fg="OFF"><rb>桐壺</rb><rt>きりつぼ</rt></ruby></div1.title>`
- `<filename id="p27" />`
- `</div1>`
- `<div1>`
- `<div1.title>きりつぼ</div1.title>`
- `<lineno pos="L" id="15"> izzurenoimi<ruby fg="OFF"><rb>時</rb><rt>とき</rt></ruby>にか。女御・更衣あまたさぶらひ給ひけるなかに、いと、やむ</lineno>`
- `<lineno pos="L" id="16"> gotonaki<ruby fg="OFF"><rb>際</rb><rt>きは</rt></ruby>にはあらぬが、すぐれて<ruby fg="OFF"><rb>時</rb><rt>とき</rt></ruby>めき<ruby fg="OFF"><rb>給</rb><rt>たま</rt></ruby>ふありけり。</lineno>`
- `<lineno pos="L" id="17"> hajimeyori, "wareha" to, <ruby fg="OFF"><rb>思</rb><rt>おも</rt></ruby>ひあがり給へる御かた、めざましき者にお</lineno>`
- `<lineno pos="L" id="18"> toshimesonemitamafu, onajiro, soreyori shimo<ruby fg="OFF"><rb>臆`

Step2 Example of Metadata

```
<!DOCTYPE record-list SYSTEM "dc-history">
<record-list>
  <dc-record>
    <title>鍋木家</title><title>鍋木太郎</title>
    <creator>千葉県海上郡海上町史編纂委員会</creator>
    <subject>海上町史料所在目録 第三集</subject>
    <subject>千葉県海上郡海上町史編纂委員会</subject><subject>海上町関係史</subject>
    <subject>鍋木家</subject><subject>江戸前</subject>
    <subject>下総国 香取郡 鍋木村</subject><subject>相給</subject>
    <description>海上町関係史料</description>
    <publisher>千葉県海上郡海上町史編纂委員会</publisher>
    <date>1981</date>
    <type>史料所在目録データベース</type>
    <format>S G M Lテキスト</format>
    <identifier>1201724</identifier>
    <source>nijl.ac.jp</source>
    <language>ja</language>
    <rights>千葉県海上郡海上町史編纂委員会</rights>
    <rights>国文学研究資料館</rights>
  </dc-record>
</record-list>
```

Metadata Creation

-Hierarchical Metadata Model-



Step3 : Standardized Data Retrieval

How to Link Heterogeneous Systems?

1. One Data Model

- Each institute has its own server
- Every server uses the same data model
- But, no flexibilities

2. Data Clearing House

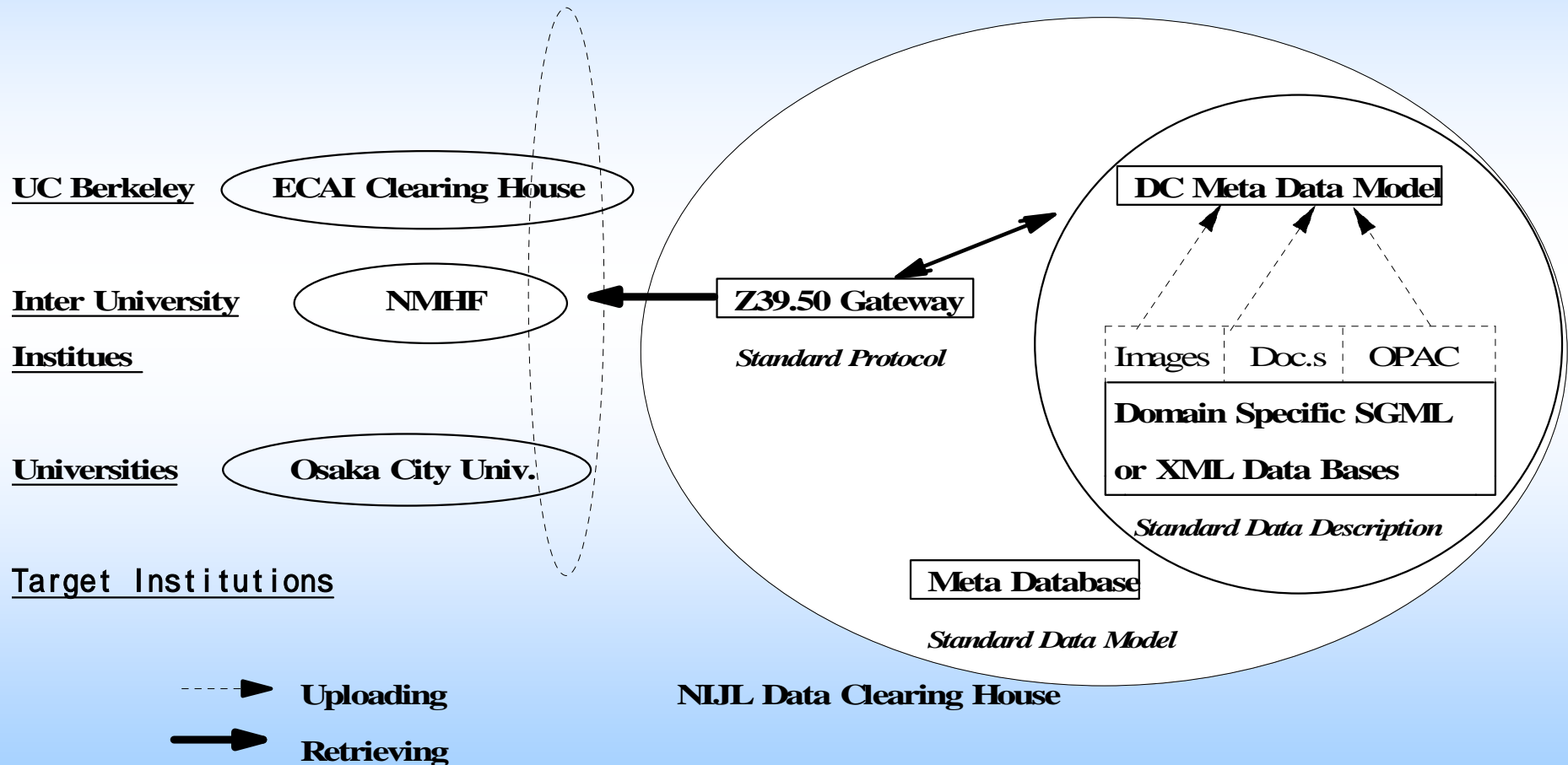
- Each institute does not need additional system
- Upload every index data to the center server
- Who maintains system, how to share costs ?

3. Metadata + Standard Protocol

- Dublin Core Metadata and Z39.50
- No center systems

How to Link Heterogeneous System?

Federation System by Dublin Core + Z39.50



Basic Components of NIJL Dana Sharing System

1. Each Database Server

- HiRDB, Bibliotheca (Hitachi Co.) etc.

2. SGML/XML tools

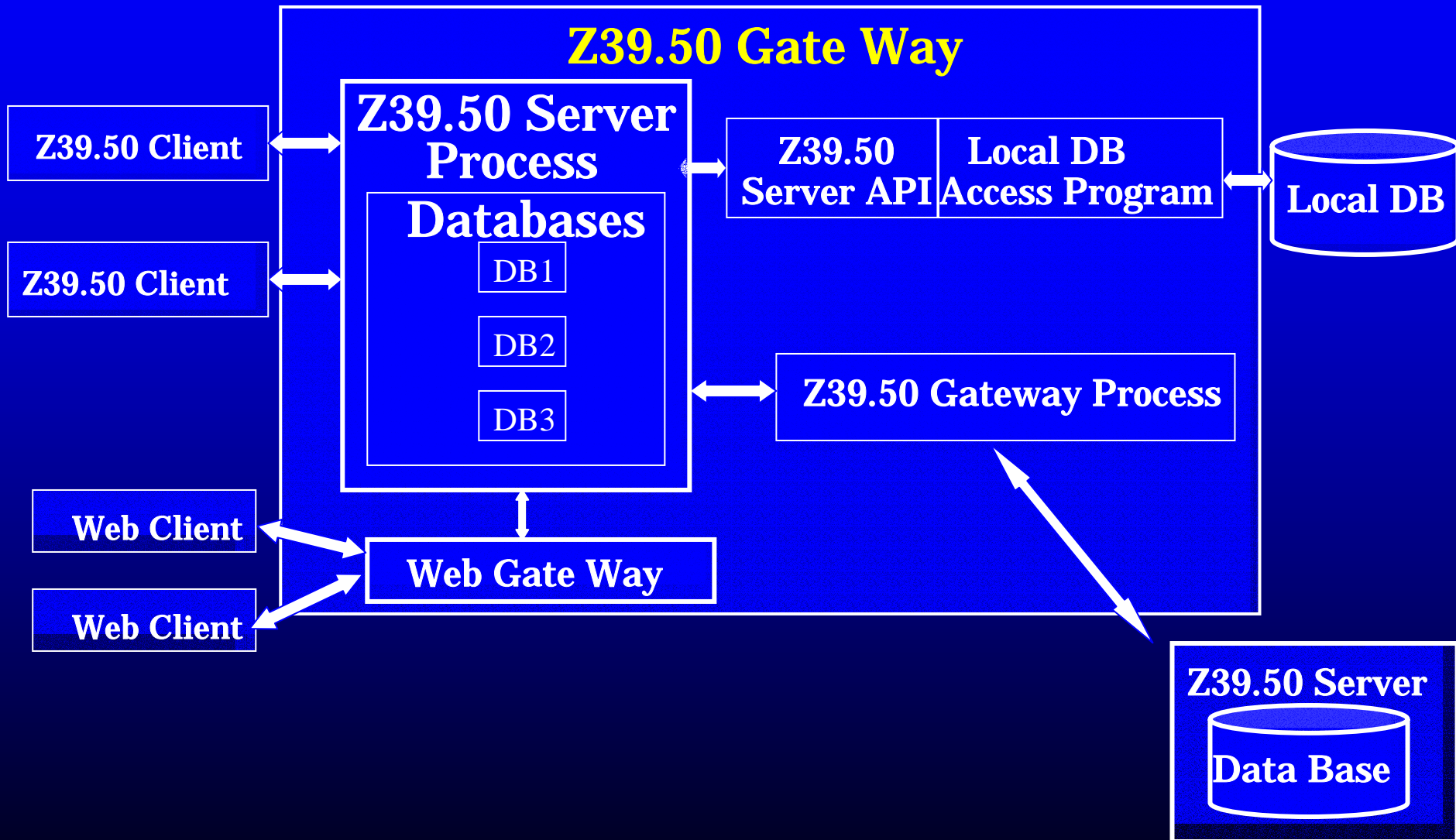
3. Dublin Core Meta Data Server

- Bibliotheca(Hitachi Co.)

4. Z39.50

- **Z39.50 Server Process**
- **Z39.50 Web Gateway**
- **Z39.50 Gateway Process**

Structure of NIJL Z39.50 System



Participating Institutes

- **The Graduate University for Advanced Studies**
 - **National Institute of Japanese Literature**
 - *National Museum of Ethnology*
 - **International Research Center for Japanese Studies**
 - National Museum of Japanese History
 - The National Institute of Informatics
 - Research Institute for Humanity and Nature
- **The Historiographical Institute, The University of Tokyo**
- **Osaka City University**
- *Keio University*
- ECAI Clearing House
- (Institute of South East Studies, Kyoto University)
- (University of Shimane)

Further Approach

- Next Z39.50 -

- **WEB Oriented**

- Remote Procedure Call
- Portability

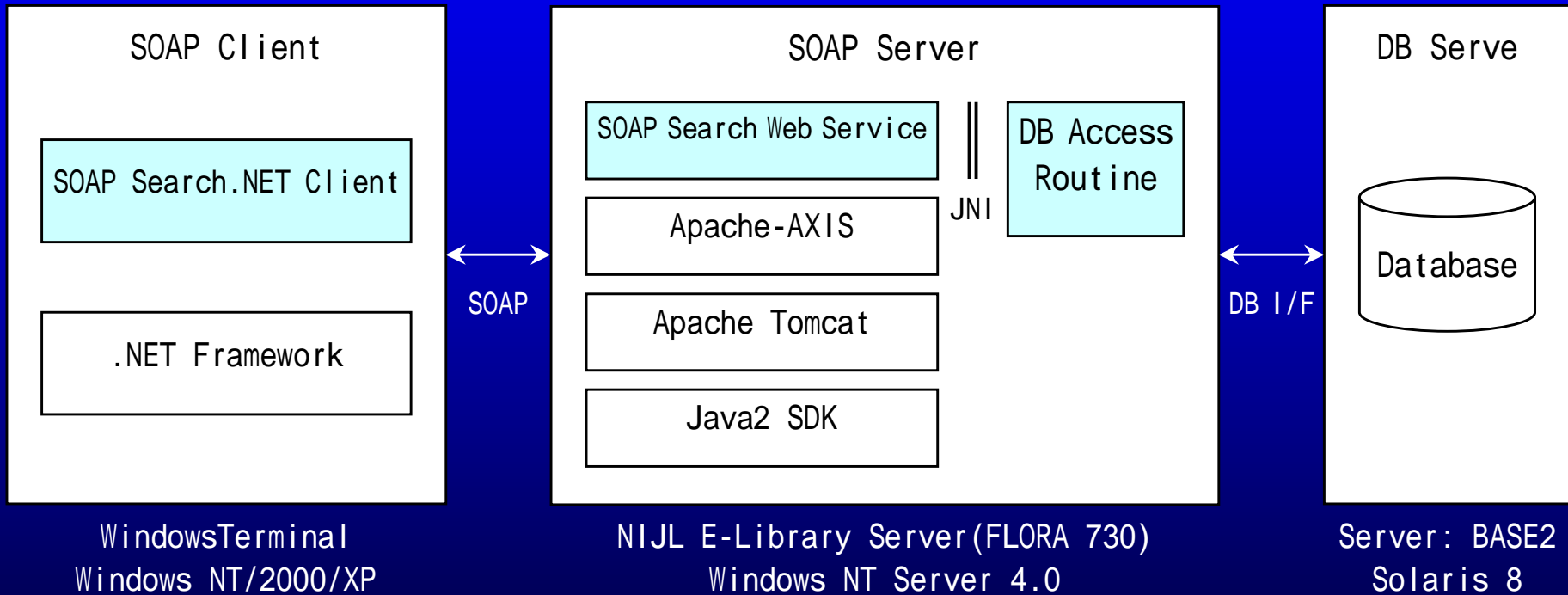
- **Light Protocol**

- Only for Data Retrieval

- **Introducing SOAP** (Simple Object Access Protocol)

- How to Treat ASN.1 ?

SOAP System



Example of Soap Message (Sender)

```
<Search>  
  <Page> Page Number </Page>  
  <DbName> Database Name </DbName>  
  <Reference>  
    <Seq>  
      <Field> Data Filed Name </Field>  
      <Contents> Key Word </Contents>  
    </Seq>  
  </Reference>  
</Search>
```

Example of Soap Message (Receiver)



```
<Result>
  <Search>
    <DbName> Database Name </DbName>
    <Reference>
      <Seq>
        <Field>Field Name</Field> <Contents>Key Word</Contents>
      </Seq>
    </Reference>
  </Serch>
  <Info> <Count> Hit Counts </Count> </Info>
  <Data>
    <Rec>
      <ItemNo> Item Number </ItemNo> <Title> Title </Title>
      <Contents> Contents of Result</Contents>
    </Rec>
  </Data>
</Result>
```

Thank You for your Kind Attention

Related URL

National Institute of Japanese Literature

<http://www.nijl.ac.jp/>

Collaboration Project: <http://world.nijl.ac.jp/~kiban-s/>

ECAI (Electronic Cultural Atlas Initiative)

<http://ecai.org>

PNC (Pacific Neighborhood Consortium)

<http://pnc-ecai.oiu.ac.jp>

Resource Sharing Group

<http://www.nijl.ac.jp/~kiban-s>

[/ResourceSharing/MeetingOnResourceSharing.htm](#)

Contact E-mail: hara@nijl.ac.jp