



# Open Platform for Academic Humanities Data

**Shoichiro HARA**

**Center for Integrated Area Studies (CIAS), Kyoto University, Japan**  
**[shara@cias.kyoto-u.ac.jp](mailto:shara@cias.kyoto-u.ac.jp)**

# Road Map of CIAS Research & Development

## ● Phase 1: Search by “Who and What”

- ✓ Digitization
- ✓ Metadata Design
- ✓ Databases
- ✓ Resource Sharing/Integratioin

## ● Phase 2: Analysis by “When and Where”

- ✓ Description of spatiotemporal attributes
- ✓ Visualizing data in spatiotemporal context
- ✓ Analysis of contents by spatiotemporal attributes
- ✓ Spatiotemporal model and tools
  - Overlay variety of maps, images, calendars etc.
  - Visualization, simulation, data mining etc.

## ● Phase 3: Discovery by Ontology

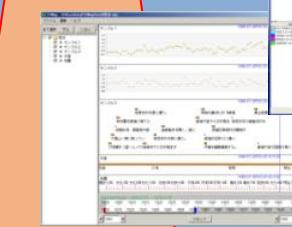
- ✓ Linking everything
- ✓ Knowledge management
- ✓ Knowledge discoveries

Database Systems



MyDatabase  
Resource Sharing System

Spatiotemporal Tools



HuMap／HuTime

年	月	日	時	分	秒	緯度	経度	高さ	状況
1999	03	15	12	30	00	35.7	139.7	1000	晴
1999	03	15	12	30	00	35.7	139.7	1000	晴
1999	03	15	12	30	00	35.7	139.7	1000	晴
1999	03	15	12	30	00	35.7	139.7	1000	晴
1999	03	15	12	30	00	35.7	139.7	1000	晴
1999	03	15	12	30	00	35.7	139.7	1000	晴
1999	03	15	12	30	00	35.7	139.7	1000	晴
1999	03	15	12	30	00	35.7	139.7	1000	晴
1999	03	15	12	30	00	35.7	139.7	1000	晴

Gazetteers,  
Chronological Gazetteers

Knowledge  
Discoveries

RDF Repositories,  
SPARQL End Point  
Text Mining and Deep Learning

# Heterogeneous Metadata

- Database is the basis of researches, BUT ... -

	Libraries, Archives	Researches
Target	Public	Individual / Research Group
Object	Public / General	Research / Specific
Collection Organization	Institutional	Individual / Research Group
Variety	Large	Large
Collection Policy	Consistent	Inconsistent / Changeable
Collection	Whole	Parts
Size	Large	Small
Metadata	Standard(generic) / Large / Complex	Heterogeneous(Specific) / Small / Simple
Usage	Simple	Complex / Inconsistent
Durability (life time)	Long	Short

## ● Challenges of CIAS

- ✓ Durable , Interoperable and Flexible Repository for Heterogeneous Datasets
- ✓ Key Technologies: Metadata + XML + HTTP + *Ontology*
  1. MyDatabase to develop databases
  2. Resource Sharing System to link heterogeneous databases
  3. REST API to realize flexible database links and usage

# Coping with Heterogeneous Metadata and Databases

## MyDatabase: Server Function for Users (Researchers) to Build Heterogeneous Databases

- **Durable Database System**

- ✓ Simple Functions ⇒ Minimum Functions
  - **Data Portability (XML)**
  - Basic retrieval functions
  - Basic GUI

- **Simple Operation**

- ✓ Simple Configuration (Minimum parameters)
- ✓ GUI

- **Minimum Constraints on Data Structure**

- ✓ **Simple Data Type (String)**
- ✓ Key field (table type) / Well-formed XML
- ✓ Free from DD/DTD(Schema)
  - CSV/TSV data: first normal form (relational data model)
  - XML data: well-formed XML document

# MyDatabase (Overview)

**InfoLib Integrated multimedia search system**

こんちは 原さん[ログアウト]

グループメニュー > サンプル原 > 歴史地震データベースサンプル1

登録データ管理

現在の検索対象データ データの取り込み

Upload

Configuration

処理対象ファイル名/形式/サイズ  
HistoricalSeismicData\_UTF8\_UPLOAD.csv CSV 347KB

合は 登録データダウンロード

検索画面設定

ヘッダ情報 フォルダ名 背景画像 配色 検索画面 一覧画面 詳細表示

バナーの表示 バナーを表示する

検索メニューの表示 検索メニューを表示する

画面タイトル 日 指定しない 英 指定しない

見出し 四 表示する 表示しない

見出しの文字サイズ 中

詳細表示の表にも表示する

空項目の表示 値の無い項目も表に表示する

コード切り替え方法 次/前ボタンを表示する ページ番号を表示する

画像の表示 全ての画像を1画面に表示する ページを切り替えて表示

画像の高さ ピクセル(10-1000)

画像のサムネイル表示 表示しない

サムネイル画像の高さ ピクセル(10-100)

サムネイル表示時の1行の画像表示件数 5件

Building

Plot count: 500 Displayed items: 500

表示対象: 1198件

Google 地図データ ©2012 GIS Innovatisa, Google, Kingway, MapKing, Mapabc, SK M&C, Tele Atlas, ZENRIN - 利用権限

Open

Data

A	B	C	D	E	F	G	H	I	J	K	L	M
ID	地震名	地震ID	史料名	レコード番号	レコードID	和暦	西暦	記載地名	現在地名	LAT	LON	ノート
1	貞元元年六月十八日	9760722	[日本紀略]	22	1151	貞元元年六月十八日癸丑	9760722	山城国／京都府／京都／上京区／小山町		35.015	135.7452778	千本丸太町交差点、近くに大極殿基壇跡の石碑(平成6年)あり
2	貞元元年六月十八日	9760722	[日本紀略]	22	1151	貞元元年六月十八日癸丑	9760722	山城国／京都府／京都／上京区／	2E-C1E	135.7452778	千本丸太町交差点、近くに朝堂院(=八省院)跡のブレーあり	
3	貞元元年六月十八日	9760722	[日本紀略]	22	1151	貞元元年六月十八日癸丑	9760722	都／東寺	市／南区／九条町	34.9777778	135.7494444	平安宮豊樂殿跡
4	貞元元年六月十八日	9760722	[日本紀略]	22	1151	貞元元年六月十八日癸丑	9760722	山城国／京都府／京都／西寺	市／南区／唐橋平垣町	34.9786111	135.7386111	左は東寺(救王護國寺)の現在地
5	貞元元年六月十八日	9760722	[日本紀略]	22	1151	貞元元年六月十八日癸丑	9760722	山城国／京都府／京都／清水寺	市／東山区／清水一丁目	34.99138689	135.7869444	左は清水寺の現在地
6	貞元元年六月十八日	9760722	[日本紀略]	22	1151	貞元元年六月十八日癸丑	9760722	山城国／京都府／円覚寺	市／上京区／柳風呂町	35.0265	135.7694444	左は円覚寺の現在地
7	貞元元年六月十八日	9760722	[日本紀略]	22	1151	貞元元年六月十八日癸丑	9760722					
8	貞元元年六月十八日	9760722	[日本紀略]	22	1151	貞元元年六月十八日癸丑	9760722					

Materials

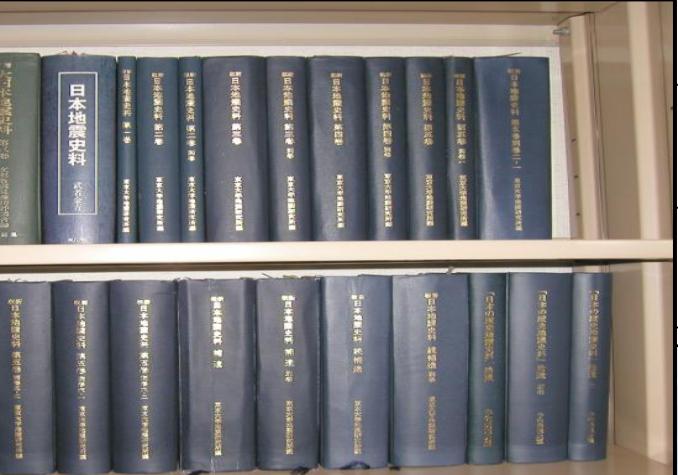
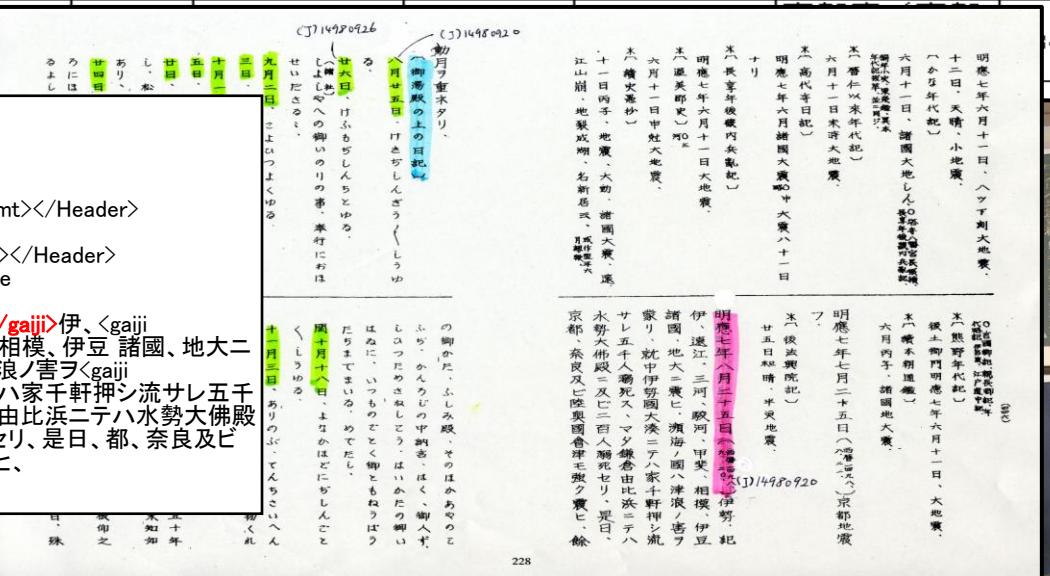
# MyDatabase (*cont. Materials*)

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	ID	地震名	地震ID	史料名	レコード番号	レコードID	和暦	西暦	記載地名	現在地名	LAT	LON	ノート
2	1	貞元元年六月十八日	9760722	[日本紀略]	22	1151	貞元元年六月十八日癸丑	9760722	山城国／京都／宮城	京都府／京都 市／上京区／小山町	35.015	135.7452778	千本丸太町交差点 近くに大極殿基壇跡の石碑(平成6年)あり
3	2	貞元元年六月十八日	9760722	[日本紀略]	22	1151	貞元元年六月十八日癸丑	9760722	山城国／京都／宮城／八省院	京都府／京都 市／上京区／小山町	35.015	135.7452778	千本丸太町交差点 近くに朝堂院(＝八省院)跡のブレートあり
4	3	貞元元年六月十八日	9760722	[日本紀略]	22	1151	貞元元年六月十八日癸丑	9760722	山城国／京都／宮城／豊樂院	京都府／京都 市／中京区／聚楽廻西町	35.015	135.7452778	平安宮豊樂殿跡
5	4	貞元元年六月十八日	9760722	[日本紀略]							4.97777778	135.7494444	左は東寺(救王護國寺)の現在地

```
<?xml version="1.0" encoding="Shift_JIS"?>
<?xml-stylesheet type="text/xsl" href="./ClassicEarthquake-Ext.xsl"?>
<!DOCTYPE ClassicEarthquake SYSTEM
      "/>ClassicEarthquakeSimple.xsd><?ClassicEarthquake?
```

./ClassicEarthquakeSimple\_ver3.dtd]]><ClassicEarthquake>  
<Volume vol="ZOTEI"><Header><titleStmt>増訂大日本地震史料</titleStmt></Header>  
<Earthquake page="228">  
  <Header><titleStmt>明應七年八月二十五日(西暦 1498.9.20)</titleStmt></Header>  
  <E.ID>14980920</E.ID><J.Date>明應七年八月二十五日</J.Date><S.Date  
    " value=""/><J.Date>14980920</J.Date>

type="Gregorian">14980920



# MyDatabase (cont. Data Upload and Configurations)

**InfoLib** Integrated multimedia search system

こんにちは システム管理者さん [ログアウト]

グループメニュー > 地域研究統合情報センター > 『貝葉文書にみる民族間関係』データベース

登録データ管理

現在の検索対象データ データの取り込み

■ 現在の検索対象データ

登録開始時間	登録完了時間	種別	処理対象ファイル名/形式/サイズ
2011-08-12 15:56:30	2011-08-12 15:56:32	新規登録	bailanlist.txt CSV 21KB

現在の検索データファイルをダウンロードする場合は [登録データダウンロード](#) を押してください。

検索対象データを前回のデータに戻す場合は [検索対象データを前回のデータに戻す](#) を押してください。

ヘルプ

ヘッダ画像2

Integrated multimedia search system

こんにちは システム管理者さん [ログアウト]

地域研究統合情報センター > 『貝葉文書にみる民族間関係』データベース

背景  
背景画像- 配色  
検索画面  
一覧画面  
詳細表示

?

baiyou-logo.jpg  削除します  
※ 更新時はファイルを選択してください。(1MBまで) 参照... 左寄せ ▾

baiyou-logo.jpg  削除します  
※ 更新時はファイルを選択してください。(1MBまで) 参照... 左寄せ ▾

日  
英  
ク  
日  
英  
日  
英  
日  
英  
日  
英  
日  
英  
日  
英

データ更新日を表示する。  データ件数を表示する。  
 データベース名・提供者を表示する。  カテゴリ、ブラウズのリンクを表示する。  
 データ更新日を表示する。  データ件数を表示する。  
 データベース名・提供者を表示する。  カテゴリ、ブラウズのリンクを表示する。

ヘッダ画像2

# MyDatabase (*cont. Open*)

標準検索 DC検索 時空間検索

矩形情報を矩形で表示 すべてマーカーで表示 マーカー+矩形をあわせて表示

表示対象: 1198件

プロット数: 500 表示件数 500

No	Title
1	永長元年十一月二十四日
2	永長元年十一月二十四日
3	永長元年十一月二十四日
4	永長元年十一月二十四日
5	永長元年十一月二十四日
6	永長元年十一月二十四日
7	永長元年十一月二十四日
8	永長元年十一月二十四日
9	永長元年十一月二十四日
10	永長元年十一月二十四日
11	永長元年十一月二十四日
12	永長元年十一月二十四日
13	永長元年十一月二十四日
14	永長元年十一月二十四日
15	永長元年十一月二十四日
16	永長元年十一月二十四日

■各項目にキーワードを入力して「検索」ボタンをクリックしてください。 録易検索 >

全ての項目:

ID:

地震名: 貞元元年六月十八日

地震ID: 指定なし

史料名: 天正十三年十一月二十九日

和暦: 慶長元年閏七月十三日

西暦: 慶長元年閏七月十二日

和暦: 文治元年七月九日

西暦: 明應七年八月二十五日

和暦: 正中二年十月二十一日

西暦: 正平十六年六月廿四日

記載地名: 永長元年十一月二十四日

LAT: 貞元元年六月十八日

LON:

表示件数: 20

検索 クリア

表示対象: 1198件

高さ変更 広げる 狹める

Timeline SIMILE

1300 1400 1500 1600 1700 1800 1900 2000 2100 2200 2300 2400 2500 2600 2700 2800 2900 3000

1BC 1000 1000 2000 3000

# MyDatabase API Application Example



# Resource Sharing System (RSS)

- **Resource Sharing System (RSS)**

- ✓ Resource Sharing System is a framework to retrieve various databases on the Internet seamlessly
- ✓ Each Database: has its own data structure in accordance with its domain specific data model
- ✓ Seamless: means that users can retrieve every database on the Internet by one operation without conscious of record structures, retrieval operations, database locations, and medias

- **Applying Some Standards**

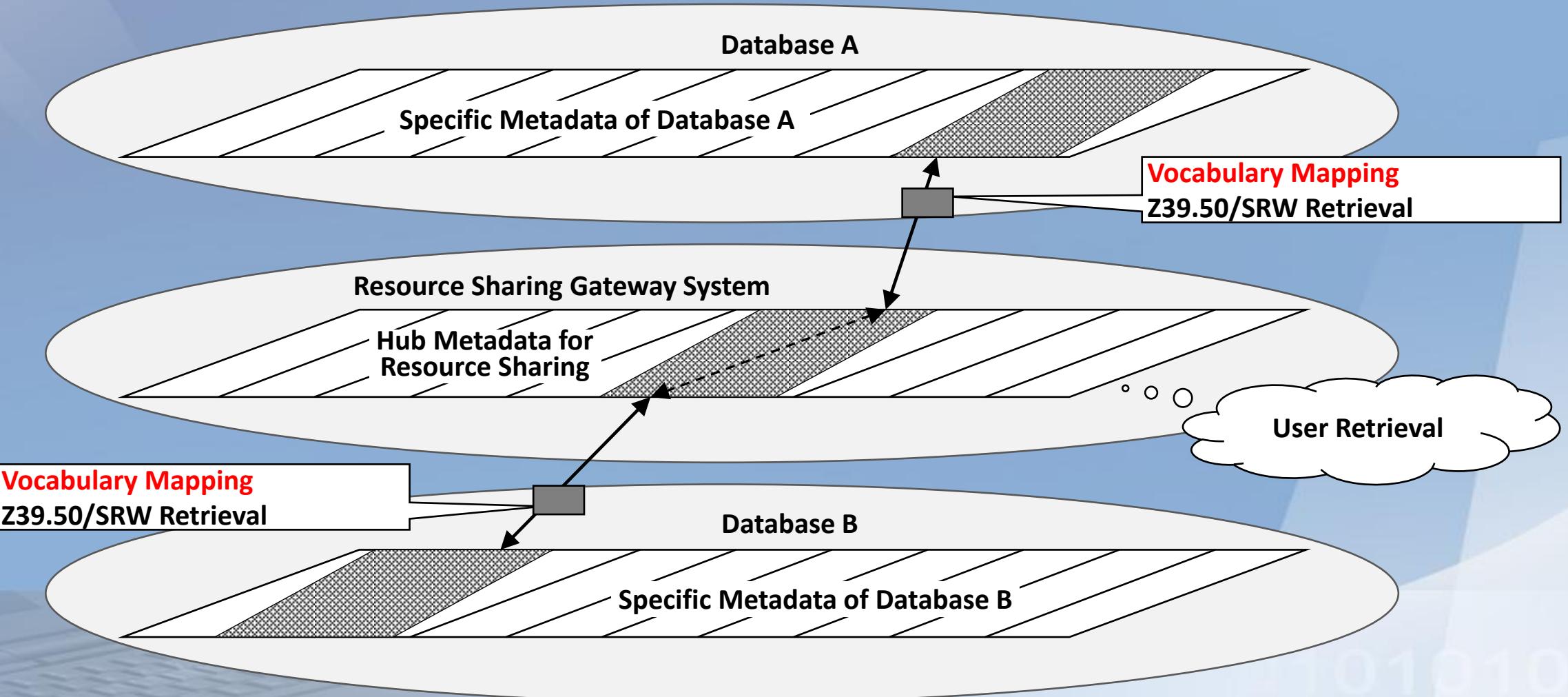
- ✓ Database (Portability)
- ✓ Data structure (Standard Metadata)
- ✓ Retrieval (Standard Information Retrieval)

- **Achievement of CIAS**

- ✓ CIAS(17), CSEAS(5), RIHN(5), NMJH(19), OPAC(5)

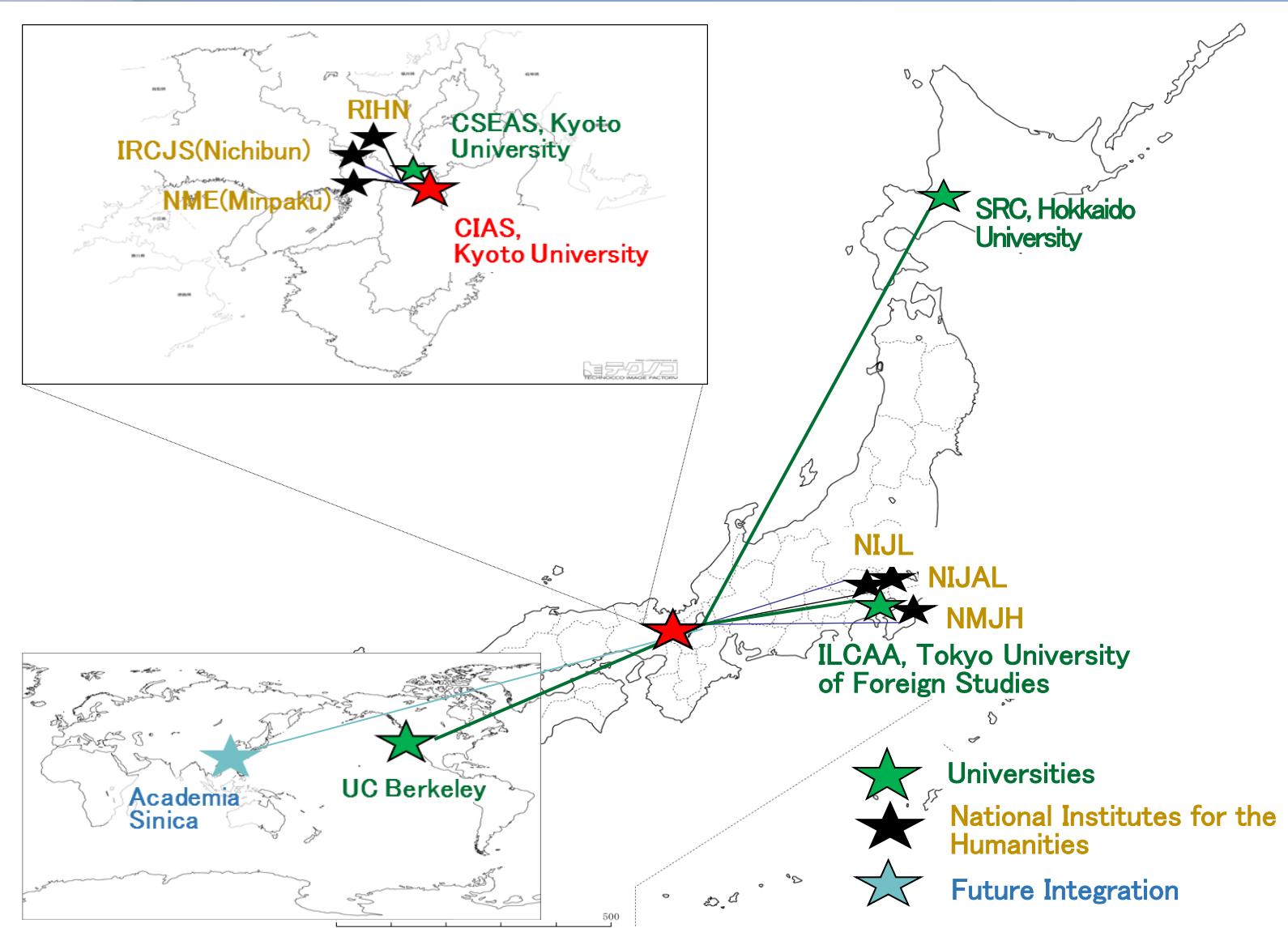
# Resource Sharing System (cont. Structure)

Resource Sharing Frontend System



# Past Development for Linking Data

## - Resource Sharing System (cont. Present Status) -



# Problems and New Research & Development

## 1. Present Resource Sharing System is not Flexible to Link Databases

- ⇒ Flexible links between university databases and cyberspace to create large-scale knowledge databases
  - ✓ data model, linked data, URI, ontology etc.
  - ✓ Text mining, natural language processing, text understanding etc.

## 2. Present Resource Sharing System is Impossible Automatically to Develop Links into Cyber Space

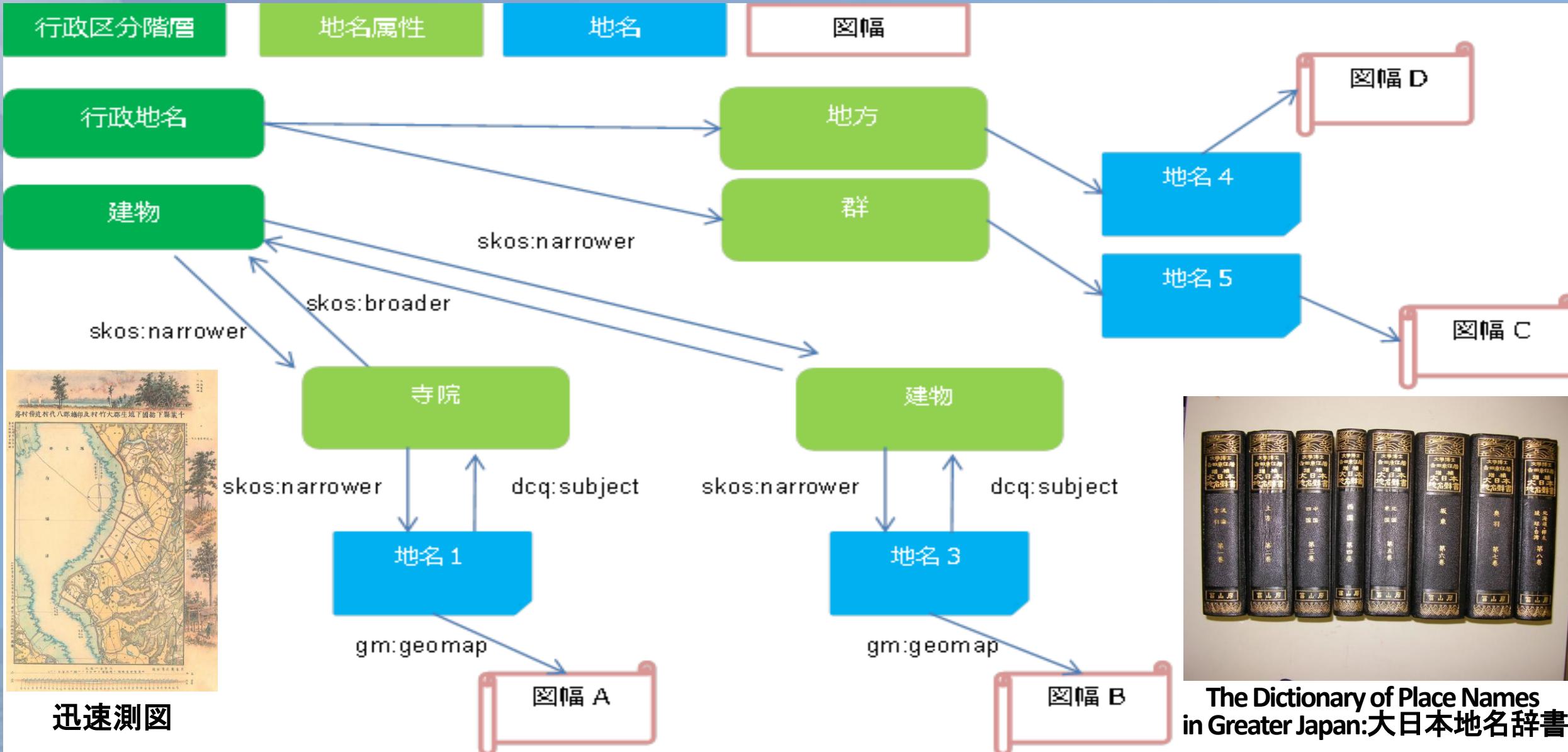
- ⇒ Development of applications to discover useful hints/knowledge for problem from large-scale databases
  - ✓ Intelligent search engine, Ontology etc.

## 3. Lack of Best Practices for Digital Humanities

- ⇒ Conducting fusion research of social science and information science in "Trans Boarder Studies on Symbiosis and Crisis"
  - ✓ visualization, anomaly detection, change detection

# RDF Preliminary Development 1

- CIAS & RIHN: Historical Gazetteer Database in Japan (Model) -



# RDF Preliminary Development 1

- CIAS & RIHN: Historical Gazetteer Database in Japan (Cont. SPARQL Endpoint) -

PREFIX rdfs: <<http://www.w3.org/2000/01/rdf-schema#>>  
PREFIX gzt: <<http://supercluster.cias.kyoto-u.ac.jp/gzt/elements/1.0/>>

```
select distinct ?place ?name
  from <http://supercluster.cias.kyoto-u.ac.jp/rdf/placename>
where  {
    { ?s rdfs:label "相国寺" }
    { ?s gzt:country ?country }
    { ?place gzt:country ?country }
    { ?place rdfs:label ?name }
}
```

100 / 583 (表示件数/ヒット件数)

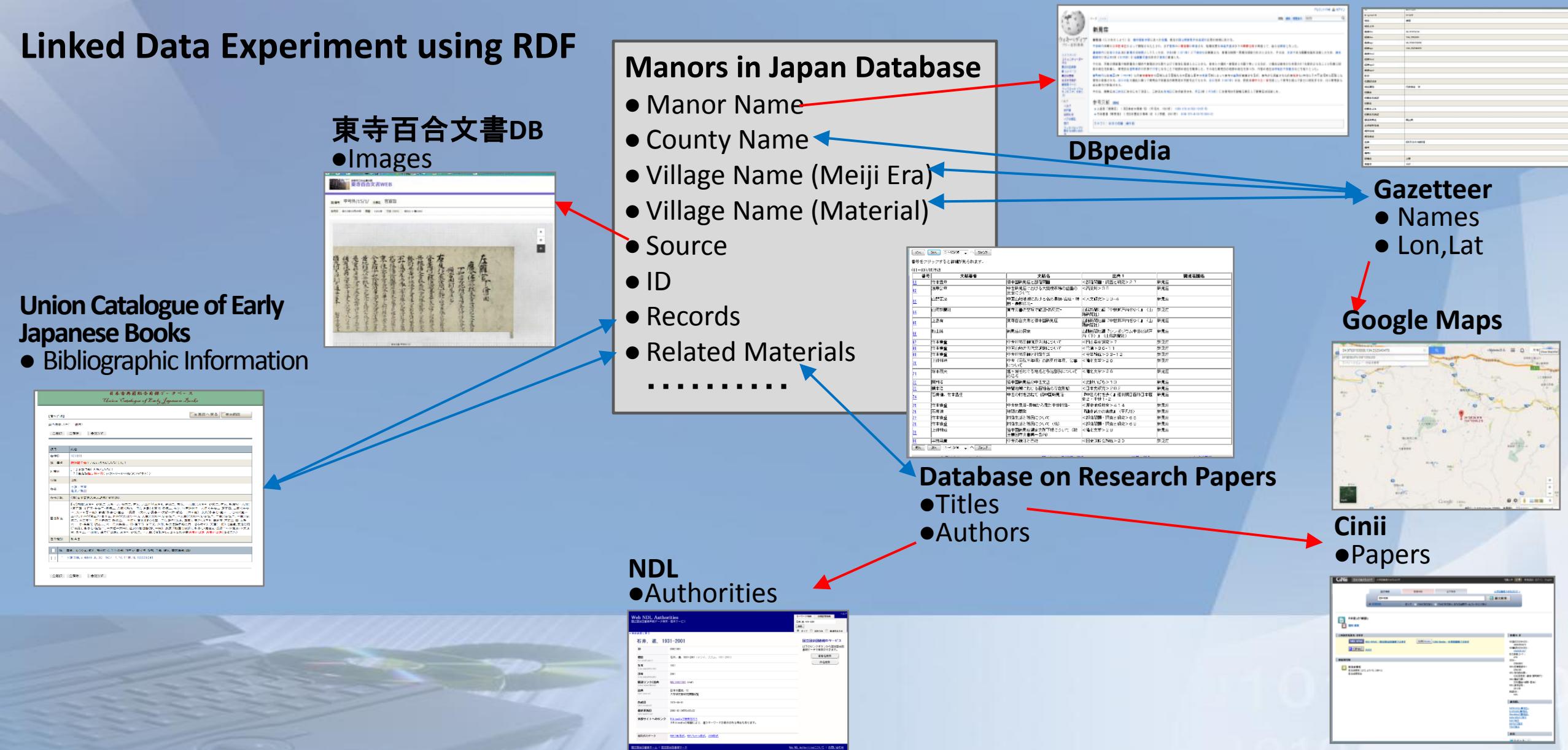
ダウンロード: RDF/XML形式 N3/Turtle形式 CSV形式

<http://supercluster.cias.kyoto-u.ac.jp/id/placename/10000017> 相河  
<http://supercluster.cias.kyoto-u.ac.jp/id/placename/10048860> 蓮華山  
<http://supercluster.cias.kyoto-u.ac.jp/id/placename/10039848> 福島潟  
<http://supercluster.cias.kyoto-u.ac.jp/id/placename/10011812> 鏡沖  
<http://supercluster.cias.kyoto-u.ac.jp/id/placename/10003401> 池之端  
...

# Linked Open Data Preliminary Development 2

## - CIAS & NIHU: *Manors in Japan Database (Model)* -

# Linked Data Experiment using RDF

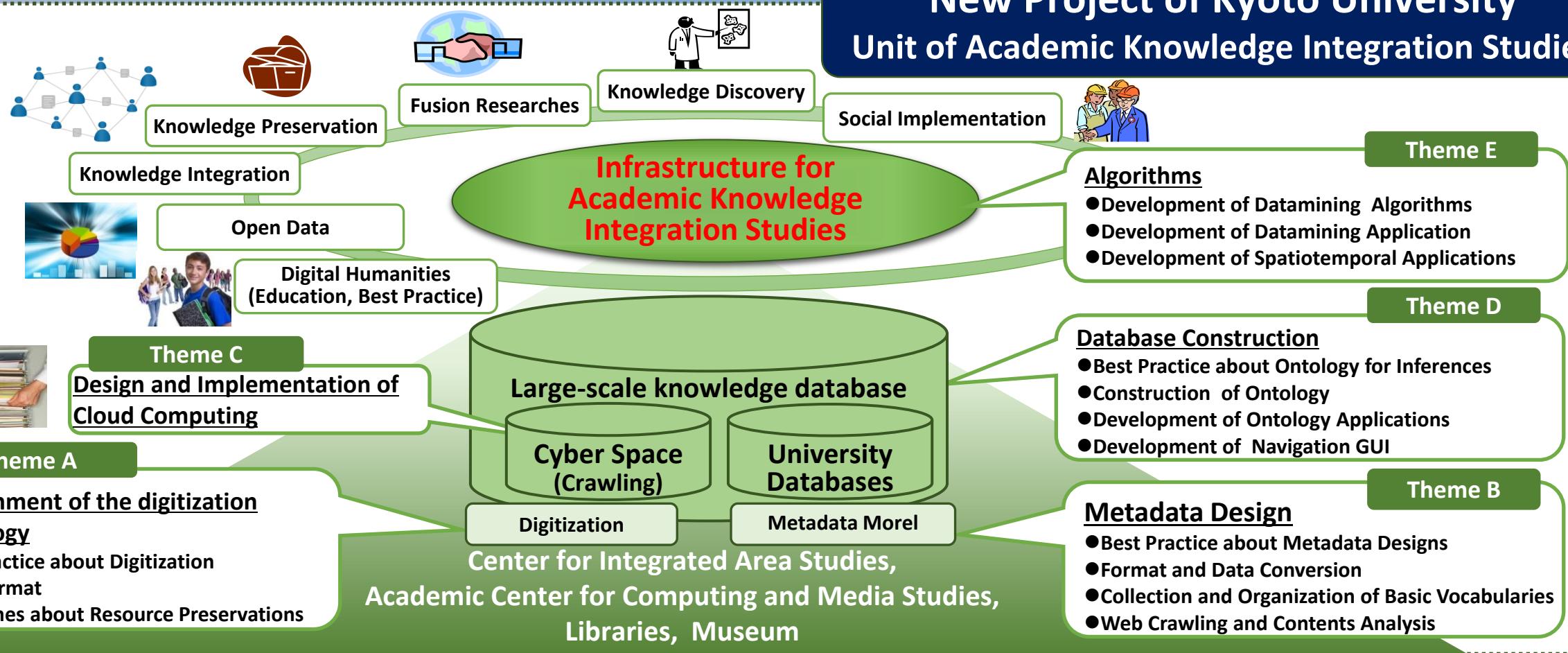


# New Project of Kyoto University

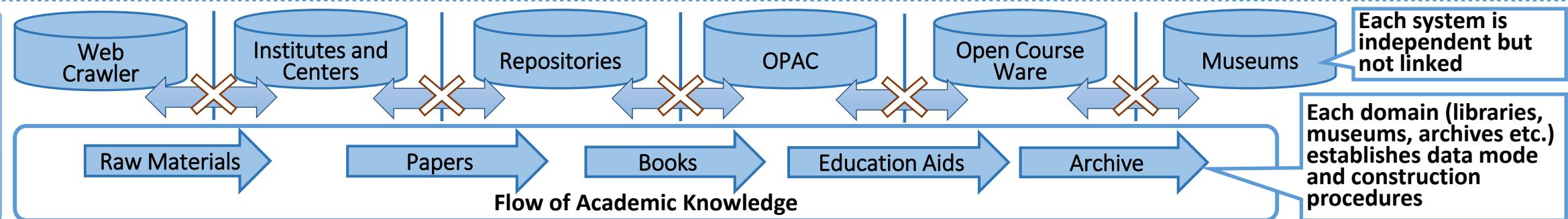
## Unit of Academic Knowledge Integration Studies



### New Objects

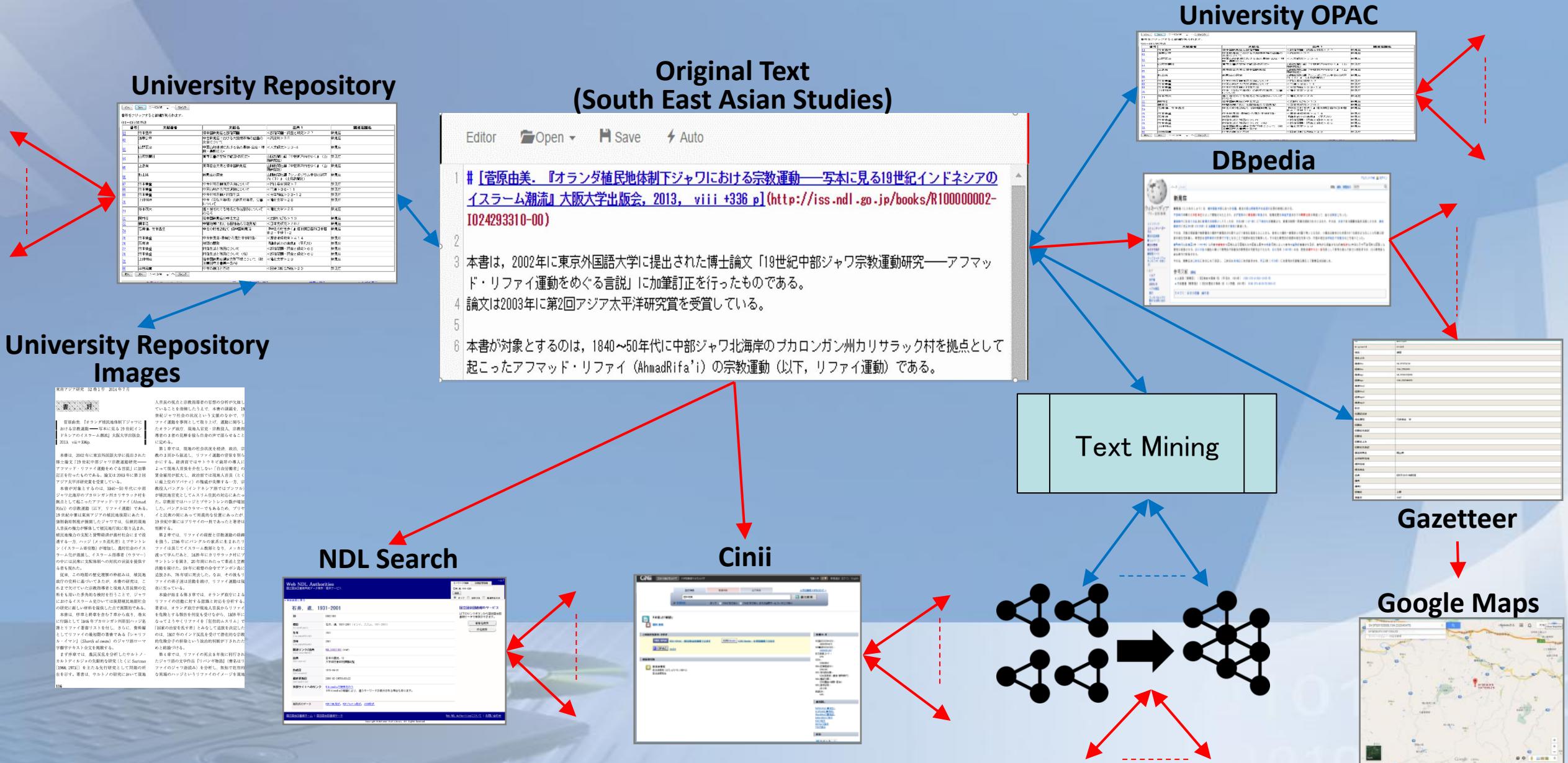


### Back Ground



Information infrastructure to integrate, open and use large-scale academic databases have been not developed

# Image of the New Project



# Conclusion

1. MyDatabase and Present Resource Sharing System have been developed to Organize and Integrate Heterogeneous Research Resources
2. A New Platform is under Construction flexibly to Link Data on Web by applying Web Sematic Technology (RDF and related techniques)
3. New Platform are Expected to Contribute to Area Studies and Digital Humanities
4. Grid and clouds are Expected to Realize Crisp LOD Processing

