京 都 大 学 防 災 研 究 所 年 報 第47号 C 平成16年3月
Annuals of Disas. Prev. Res. Inst., Kyoto Univ., No. 47 C, 2004

Development of Cross-Media Database for Sharing Disaster Information

Tomohiro KUGAI *, Yoshiaki KAWATA and Haruo HAYASHI

* COE Researcher, DPRI, Kyoto University

Synopsis

This document describes the system design and implementation on Cross Media Database. It is a research support information infrastructure (RSII) being developed for researchers. It is aimed to provide integration and interoperability of various information related to research activities. It can catalog, locate and retrieve a wide range of information sources regardless of their format. It will allow for the growth and evolution of the system over time to support multi-disciplinary professionals and scientific teams, and other stakeholders all over the world.

Keywords: Database; Multi-media; Multi-relationship

1. Introduction

Cross Media Database is a research support information infrastructure (RSII) being developed for researchers. It aimed to catalog, locate, and retrieve a wide range of research-support information sources regardless of their format: spreadsheet data, reports, geospatial databases, multi-resource type including video, still images, and audio, web pages and data about people, organizations, and events. It will provide integration and interoperability of a broad range of information related to various research activities, and will allow for the growth and evolution of the system over time to support multi-disciplinary professionals and scientific teams, and other stakeholders all over the world.

Cross Media Database provides a framework in which researchers can catalog, store, retrieve, search, browse, discover, and visualize information effectively, and examine a broad range of information related to their research topics efficiently. Cross Media Database is based on the principle that resources have “relationships” to other resources. Additionally, Cross Media Database utilizes spatial and temporal display interface to present the distribution of information geographically as well as chronologically. It also provides detailed information about resources by storing metadata and digital files. The metadata formats used in Cross Media Database conform to various international standards allowing information to be shared or transferred between other compatible systems over the Internet.

2. Concept design of Cross Media Database

Researchers of disaster domain studies will have needs that they want to use many kinds of resource types effectively. Cross Media Database has the following characteristics in consideration of needs of researchers.

[1] Cross Media Database can store many kinds of resource types, media types (Physical objects stored data/information, e.g., Paper, Film, Compact Disc and DVD) and format types (Format of Digital data or Physical, e.g., Text, Doc and Spreadsheets) as comprehensive database.

[2] Cross Media Database makes use of metadata set as international standard and integrate multi resource types by using metadata.

[3] Cross Media Database defines the relationship explicitly between resource types and authorizes it.

[4] Cross Media Database has functionality that can retrieve and view to be based on Time and Geospatial.

[5] Cross Media Database has expansibility and flexibility that we can add functions and definitions in correspondence with needs of stakeholders.

2.1 Multi resource types

We defined twelve data types to store in Cross Media Database.

[1] Audio: Sound recording in and on all types of media, physical or digital.

[2] Data: Distinct pieces of information usually formatted in a special way. Data can exist in a variety of forms - as numbers or text on pieces of paper, or as bits and bytes stored in electronic memory.

[3] Document: A physical or digital entity containing all or a portion of a work or several works.

[4] Event: An occasion or activity such as meeting and conference.

[5] Geospatial: Geospatial is information that identifies the geographic location and characteristics of natural or constructed feature and boundaries on the earth. This information may be derived from, among other things, remote sensing, mapping, and surveying technologies. Information is organized as datasets that comprise a geographic information system (GIS).

[6] Image: The graphics that does not have a position coordinate represented by a photograph or a picture.

[7] Internet: A resource located on the Internet.

[8] Model: Computer modeling software whose primary function is to model a certain class of physical systems. It may include pre- and post-processing components another necessary ancillary program.

[9] Organization: Human beings making up a group or assembly, or linked by a common interest or administrative/functional structure.

[10] Person: Human, individual.

[11] Study: Funded research project aimed at the discovery and interpretation of facts, revision of accepted theories or practices in light of new facts, or practical application of new/revised theories or approaches to a problem.

[12] Video: A physical or digital entity containing all or a portion of a work or several works on film, videotape or electronic file. These would include complete commercial films and programs, compilations, trailers, newscasts, and unedited non - commercial taped footage.

2.2 Metadata

Metadata is data about data and is described the characteristic in order to identify, describe and retrieve resource and for example, the content is described creator, date and title. Cross Media Database makes use of metadata set as international standard and integrate multi resource types, multi media types and multi format types by using metadata whatever it is digital or not. Metadata and structure of Cross Media Database allows storing data for multi resource types and user can retrieve data easily. Therefore, researchers can search metadata in Cross Media Database, and distinguish it whether they need or not quickly, and identify the location where data is if data is available.

In Cross Media Database, two types of metadata element sets: “Common elements” and “Attribute Elements”. The former is common to all types of resources, and the latter depends on the each resource type. Every “Attribute Elements” has a special metadata element that is based on a format as international standard. Attributes are defined format in correspond with characteristics of resource type. For example, MARC21 is metadata format for bibliography information; Dublin Core is metadata format for Internet resource and FGDC is metadata format for Geospatial information.

- Audio: MARC 21(MARC URL)
- Data: FGDC(FGDC URL)
- Document: MARC 21
- Event: Gale Associations(Gale URL)
- Geospatial: FGDC
- Image: VRA Categories, Kodak, MARC 21(VRA URL)
- Internet: Dublin Core
- Model: UCSB work on Computer Model Metadata
- Organization: Gale Associations
- Person: Gale Associations
- Study: FGDC Biological Extensions
- Video: Dublin Core

[1] MARC21 (Machine Readable Cataloging 21) is the standard format for the representation and communication of bibliographic and related information in machine-readable form. It is developed and maintained by the United States Library of Congress (LOC).

[2] Dublin Core. The desire for interoperable metadata about World Wide Web resources spurred the formation of the Dublin Core Metadata Initiative. A list of “core” metadata elements “to provide vertically specific (or semantic) information about Web resources, much in the same way library card catalogs provide indexed information about book properties” was developed in 1995. Dublin Core metadata is used to supplement existing methods for searching and indexing Web-based metadata. The original discussions that took place at the first workshop were focused primarily on creating metadata for electronic resources however since that time the consensus among the DC community is that DC-enabled resource discovery systems can and should be used to describe both digital and "real" physical objects. Most DCMI participants are involved in large-scale archiving or cataloging projects that require the use of Dublin Core metadata to enable large collections of object “resources” to be grouped, named, classified, and indexed in a useful fashion.

[3] FGDC. Geospatial metadata gained national attention in the mid-nineties when Executive Order 12906, "Coordinating Geographic Data Acquisition and Access: The National Spatial Data Infrastructure," was signed by President William Clinton. Section of the executive order calls for the development of an online national geospatial data clearinghouse with standardized data documentation. In 1998 the Federal Geographic Data Committee drafted a national standard for geospatial metadata: the Content Standard for Digital Geospatial Metadata is commonly referred to as "FGDC".

2.3 Relationships

Researcher will have processing, unification, analyzes various data and/or information in study processes. In this process, processing or integrating some raw data will raise new data. At the same time, researches will create knowledge with data. In addition, when researchers search some information, they might want to know relationships that are "Who is author of the thesis?", "Which organization does author belong to?", "What kind of interest the author is?" or "What event or conference is held about the thesis?" In Cross Media Database, relationships are defined in advance explicitly between resource types and authorize it. User can understand relationships easily between resources and collect data/information effectively. Relationships are defined such as "Presented at" (for Event, Organization), "Component of" (for Audio, Document, Internet, Study, Video, Image), "Commissioned for" (for Event, Person, Study, Organization).

Figure 1 shows the Relationships. Document1 is a document from one publisher and also is written by person 1. The relationship is an author between document1 and person1. And person1 is also an author of document2. Moreover document 1 is a document about one topic. Another person (Person2) is an author of document3 and document3 is paper of same topic to document1. Regarding to this topic, there are some dataset and image are and user can find some data easily.

2.4 Retrieval by Topics, Time and Geospatial

In Cross Media Database, user can retrieve information/data view the results based on topics, time and geospatial. Especially, view based on time and geospatial are new way of visualization of search results. In this new function, user can identify where monitoring data is and where picture is taken with location data and map. Generally, though geospatial data has information about location, all resource type can have footprint. In addition, user can view data chronologically by using metadata such as creation date or content date. It is very useful for researches to know data by time and geospatial.

2.5 Flexibility and Extendibility of System

Cross Media Database has concept that it can grow up by needs of researchers. It means that we can add and change contents of controlled vocabulary and define new relationships by needs of researchers. So Cross Media Database has flexibility and extendibility of system.

3. Database scheme

Cross Media Database, as shown in Figure 2, consists of several tables and relationships among them.

These tables can be categorized into two groups: "Resource Elements," that stores metadata on resources, and "Relationships," that stores relationships between resources.

3.1 Resource Elements

Metadata on resources consist of two groups:

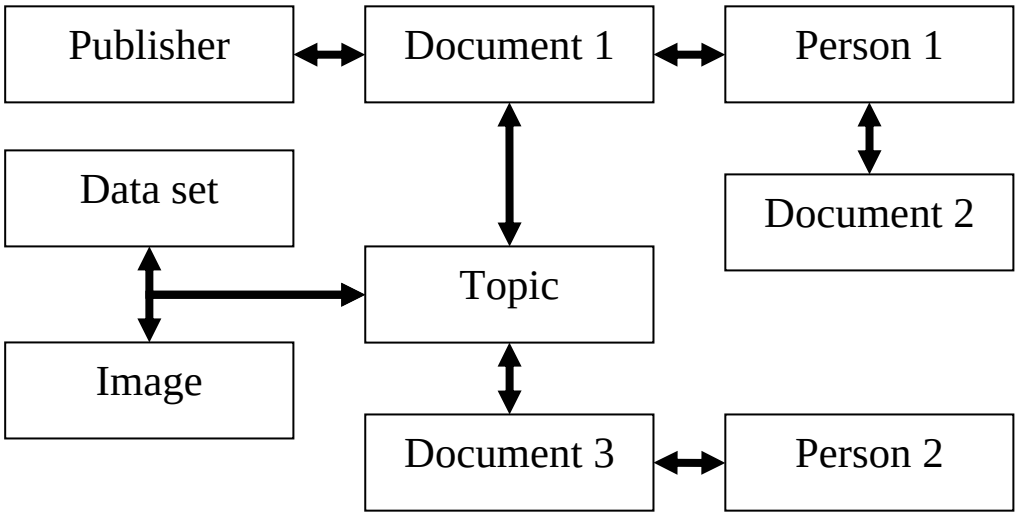

Fig. 1 Relationship

"Common Elements," that are independent of resource types, and "Attribute Elements," that are depend on each resource type.

[1] Common Elements are the element set that are stored for all types of resources: title, keyword, abstract, publishing date, location of a resource, resource type, etc. Resource Base is the core table of all in Cross Media Database. It stores some types of elements whish are recorded at most one item for each resource. The primary key in Resource Base is Resource ID, which is generated automatically when a resource data is newly inserted. It is used as a foreign key in other tables to point the resource. Values, which are unique to each resource such as title and abstract, are stored directly in the Resource Base table. On the other hand, the ID of selected choice is stored for elements which are selected from lists of choices such as Media type and Format type. Cross-reference tables are used for multi-selectable items such as languages and keywords. It stores couples of Resource ID and the ID selected item.

[2] Attribute Elements are the element sets that store the characteristics of each resource type: Volume numbers and ISBN for documents, size and number of colors for image, etc. These data are stored into tables that are prepared for every resource type. The primary key of each table is Resource ID, and it is also the foreign key.

3.2 Relationships

Cross Media Database provides the relationship retrieval. It shows a hyperlink on the result page of retrieval when one resource has a relationship to another. Users can get more information on the resource with following it. Relationships includes; "Author" between "Document" resource and "Person" resource, "Reference document" between two "Document" resources. The available type of a relationship is depending on both sides of resource types that it connects.

4. Tuning up for disaster studies

As shown above, the framework of Cross Media Database is flexible enough to cover large area, which does not only focus on disaster studies. We have to modify and configure details of Cross Media Database to fit disaster studies. It includes arranging of metadata element sets, arranging of controlled vocabularies, and communication to collaborate with outer database systems.

4.1 Controlled vocabularies

Controlled vocabularies are used for items where free inputs are not permitted for quality control, including themes, categories, format, etc. At present, we have adopted three axes of categories: "Disaster type", "Disaster life cycle type" and "Study type".

Elements of these controlled vocabularies are selected from the basic plan for disaster prevention, categorizations in academic societies, and dictionaries and glossaries on disasters. Vocabularies are easy to extend. Though we had decided them through discussions and K-J Method, it is important to refine through interviews with researchers in DPRI and other research institutes to make more useful and

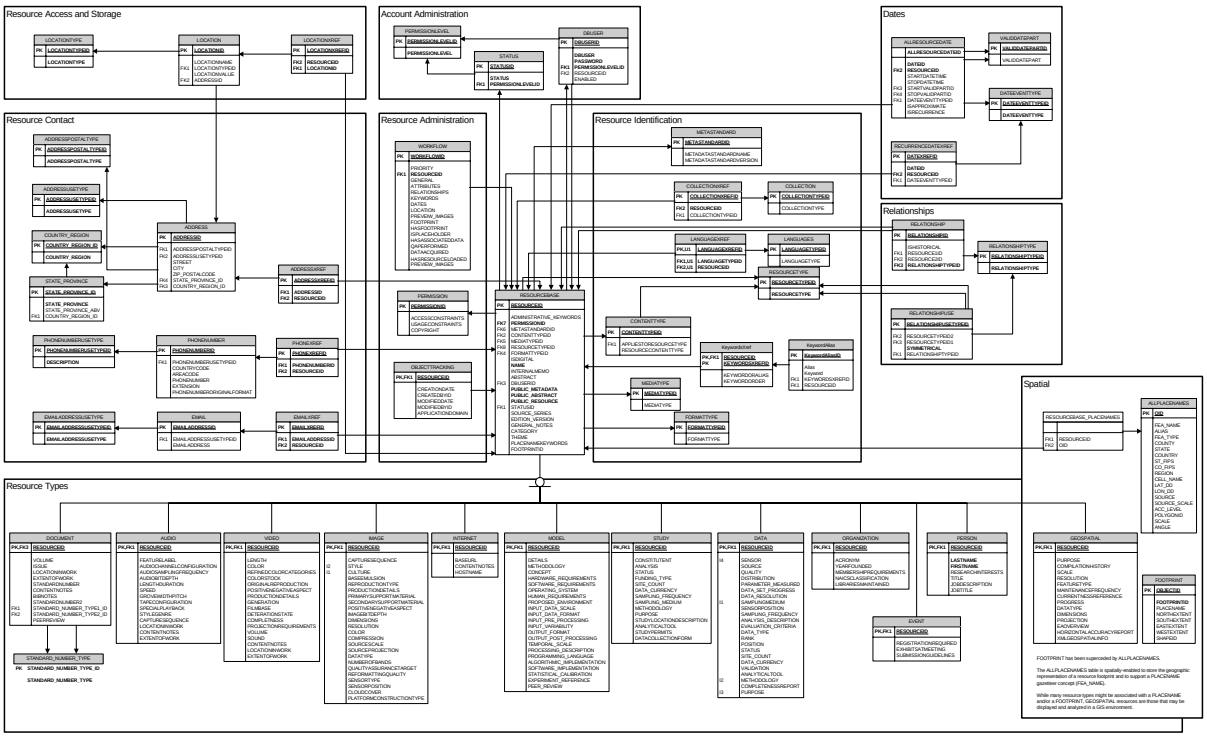

Fig. 2 Database scheme of Cross Media Database

researcher oriented.

[1] Disaster type consists of twelve types: “Earthquake”, “Atmospheric”, “Water”, “Geo-disaster”, “Volcanic”, “Snow”, “Traffic”, “Nuclear”, “Hazardous Material”, “Fire”, “Pollution” and “War and Terrorism”.

[2] Disaster Life Cycle type consists of four types: “Preparedness”, “Response”, “Recovery” and “Mitigation”.

[3] Study type consists of four types: “Observation, Management and survey”, “Theory and Mechanism”, “Information Communication” and “Information Technology”.

Catalogers can select multiple types without considering boundaries of three categories. In addition, they can select nothing if they could found no types suite for the resource. Resources should be cataloged in researcher oriented way, and retrieved in user oriented one.

Keyword

On the contrast, controlled vocabularies are not suitable for some kinds of elements. For example, “Title” and “Abstract” are usually unique to every resource. Full-text indexing is used to search these elements. Free words are used to search resources like web search engines.

In case of “Keyword”, vocabularies should be controlled for quality and thesaurus control, however it is not easy for catalogers and librarians to understand various keywords in whole study fields. In addition, authors of papers tend to give keywords not based on any standards.

4.2 Collaborate with other database systems

Cross Media Database are ready to extend to collaborate with other database systems. Cross Media Database can accept queries and transmit result on http (Hyper Text Transfer Protocol) to other systems because it has web-based search retrievals. Queries can be carried in a format of each system on CGI parameters over http, and results in Dublin Core format on XML over same protocol. At present, Database SAIGAI and DPRI Nenpo search systems are considered to our partner because these are also developed as http-based systems.

5. Conclusions

Cross Media Database aimed to be the research support information infrastructure supporting multi-media and format resources and their relationship, and multi-retrieval components. It can communicate with other database system because international standard metadata sets are supported.

For the future, to make the system more useful, we have to store a lot of records and types of resources. In addition, it is important to refine through collaboration with various disaster researchers.

Acknowledgments

This research was supported by the Japanese Ministry of Education, Culture, Sports, Science and Technology (MEXT) 21st Century COE Program for DPRI, Kyoto University (No.14219301, Program Leader: Prof. Yoshiaki Kawata).

References

Urakawa G., Yoshitomi N., Kugai T., Kawakata H., Topping K. and Hayashi H. (2003): Development of Cross-Media Database for Sharing Disaster Information, *Proc. 4th Workshop for “Comparative Study on Urban Earthquake Disaster Management”*, DRS, DPRI, Kyoto Univ.

要　旨

クロスメディアデータベースは、防災領域研究者が幅広い分野の研究を効果的に進めるための研究支援情報基盤システムである。データの種類を問わず、利用者のニーズに応える形での検索及び表示を目的とする。メタデータの項目や統制語彙、リレーションシップに関して、防災学の各分野に適合するよう検討するとともに活用事例や利用者の意見をフィードバックし、よりニーズにあったものへと拡張していくことが可能である。

キーワード：データベース，マルチメディア，リレーションシップモデル