京 都 大 学 防 災 研 究 所 年 報　第48号 C　平成17年4月
Annuals of Disas. Prev. Res. Inst., Kyoto Univ., No. 48 C, 2005

Building an Integrated Database System of Information on Disaster Hazard, Risk, and Recovery Process – Cross-Media Database (2)

Hironori KAWAKATA, Paul YOSHITOMI, Go URAKAWA, Kelly CHAN,
Hideki MATSUURA, Kenichi TATSUMI, Takeshi HARA, Munenari AGUSA,
Haruo HAYASHI, and Yoshiaki KAWATA

Synopsis

Sharing of various information and data on prediction, prevention, communication of information, history, recovery and revival processes, etc. of disaster is indispensable for successfully implementing a total risk management for disaster. In order to utilize these types of data and information, widely and effectively, we focused our research on building an integrated database system (Cross-Media Database) for disaster reduction, and a prototype has been developed using data on the 2004 Niigata Prefecture Heavy Rain Disaster and the 2004 Niigata Prefecture Chuuetsu Earthquake Disaster a case studies.

Keywords: database, disaster, Cross-Media Database (XMDB), metadata, relationship, research support information infrastructure

1. Introduction

Disaster reduction science is an interdisciplinary field composed of lots of special subjects of study, such as civil engineering, architecture, earth science, social science. The disaster and damage is controlled by combination of various factors, such as regional features, social foundation, geographical and geophysical features, and the physical mechanism of the causal natural phenomena. Therefore, in order to establish disaster reduction science and to perform the suitable total risk management about disaster, sharing of various information, data, and research results on disaster science, such as prediction, prevention, history, communication of information, and recovery and revival processes of disaster, is indispensable, as well as human communication among researchers on disaster science. However, a researcher is often difficult to use data and research results of other field than his specialty, and is difficulty to understand them. If suitable explanations for them cannot be obtained, effective cooperation is barred and misapprehension of meaning of data tends to lead the wrong results. Then, a new integrated database system, one of a research support information infrastructures, is needed that has a user-friendly interface for the researcher belonging to various disaster science fields, and can store the data of various disaster science fields saved on various formats/media. In this project, we are trying to construct a new integrated database system (Cross-Media Database) that can respond to the needs of the various field of disaster science study. This paper is the follow-up of the paper in the last year (Kawakata et al., 2004), and shows the design of a system, and the outline of Cross-Media Database (XMDB) prototype.

2. XMDB System Design

2.1 System Overview

XMDB combines features from traditional library catalogues, from those of digital libraries, and has certain characteristics of an internet search engine. The core database resembles a digital library. As in a digital library, XMDB supports full-text indexing, sometimes against various scientific thesauri that are hierarchical. It also has the capability for static and dynamic browsing of its catalogue. Furthermore, XMDB includes two additional, distinct features. XMDB supports explicitly encoded "relationships" amongst its holdings as a form of knowledge management. XMDB also includes infrastructure to record access and usage of its resources, hence providing the necessary data points for collaborative filtering. Collaborative filtering is a technology for tracking user search patterns, thus, making automated associations between information. This can provide users with recommendations of related information that people with similar interests have also discovered. Collaborative filtering is not currently implemented in our system.

XMDB began as the SSDP (Salton Sea Database Program) Cross-Media Database, at the Redlands Institute (Center of Environmental Studies, University of Redlands, California, USA), in 2001. There have been four development phases before we adopted it to manage its research holdings in disaster management at DRS (Research Center for Disaster Reduction Systems), DPRI (Disaster Prevention Research Institute), Kyoto University. The DRS version is derived from Cross-Media Phase 3-SPINE implementation. SPINE was the first end-to-end deployment of the Cross-Media system. Its objectives were to capture research holdings for the desert tortoise study at the Redlands Institute.

The Cross-Media system comprised of a database and three end user applications: Cross-Media Cataloger, a windows-based interface for data entry; Cross-Media Internet, a web-based interface for searching and display of data across the internet; and Cross-Media Explorer, a windows-based interface for searching, browsing, and spatial/temporal data exploration on the desktop. The DRS installation contains the database, cataloger, and Cross-Media Internet. It does not include the desktop Cross-Media Explorer application as all of Explorer's functions have been migrated to the web-based interface. Fig. 1 shows a stylized representation of the system.

The deployment of XMDB at DRS introduced a set of server components to the system. These components support a service architecture to access the database. This design allows further isolation of end user applications from the database. This enhancement has since been integrated back into the main code tree at the Redlands Institute.

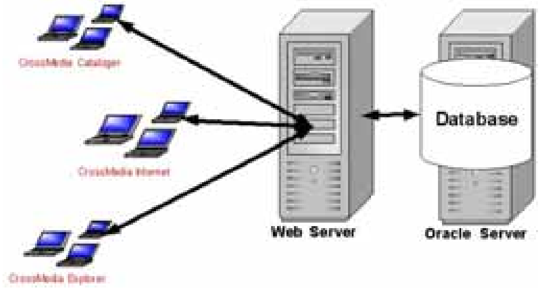

Fig. 1. A stylized representation of the XMDB system

2.2 Architecture

XMDB at DRS is implemented using Microsoft SQL Server as the database engine. The bulk of the system is developed using commercial off-the-shelf (COTS) technologies from Microsoft and ESRI (Environmental Systems Research Institute). There are occasional uses of open-source software in the form of the Apache Tomcat web application server and Perl scripts.

XMDB always has a modular design, especially with the post-Phase 3 services oriented architecture at DRS. This modular design minimizes interdependencies amongst components. Many of the components have been built by independent research teams, and many are reused from other projects.

The XMDB system at DRS is broken into three main sub-systems. These are (1) the server components, (2) the Cataloger (sometimes referred to as xmdbNG) data entry application, and (3) the xmdbSearch_Kyoto web site. All components currently work in a Windows 2000 environment (The system has been running under Windows XP Professional without modifications. The system still runs on Windows 2000 servers to take advantage of its SQL Server supports), running the dotNet 1.1 framework. The spatial portion of the database runs on ArcSDE 9.01 from ESRI. Map displays and geoprocessing tasks are performed on ArcIMS 9.01.

Fig. 2 highlights the major software building blocks in the system. The bottom three items are commercial products. The two items in the top tier represent the two end user applications custom-built for XMDB. Web services in the middle layer are highly customized services; these include customized database queries and mapping services.

2.3 End User Applications

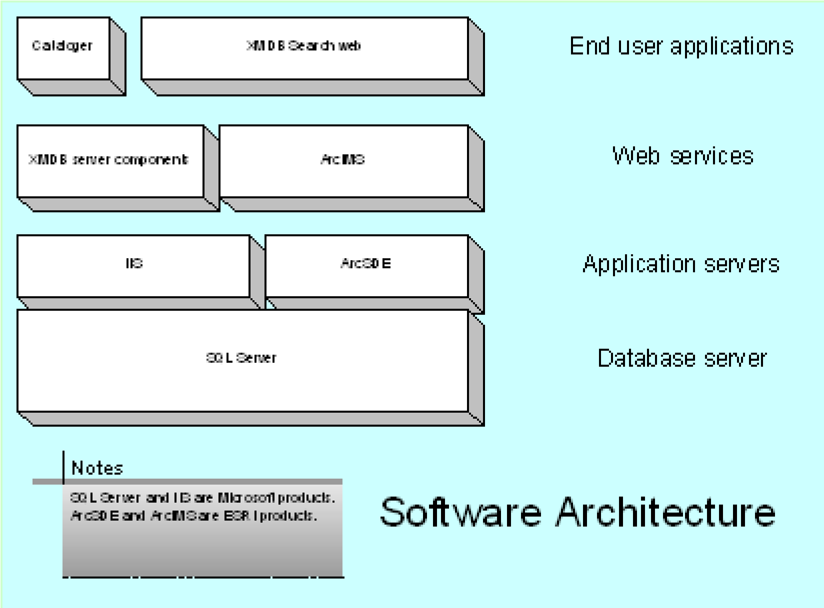

Fig. 2. The major software building blocks in the XMDB system

Most interacts with XMDB using one of the two end user applications. The web search application allows easy access to the catalogue and the holdings. Text searches are available for catalogue data, other meta-data, and full-text when documents have been captured on-line. Two non-textual searches, geographic and timeline searches, respectively, are also supported. The XMDB Search web application is described in additional details elsewhere in this report.

The Cataloger is provided to input and maintain catalogue data. It provides full access to the database. This is a conventional form-based data entry application. Presently the Cataloger does not handle spatial data.

2.4 Server Components

Six separate, but inter-dependent, components make up the server module. These six are: (1) Cross-MediaSearchGenerator, (2) dataaccesslayer, (3) remotingXMDBDal, (4) SearchAPI, (5) vbXMDBCustomArrayList, and (6) XMDBORLayer. Fig. 3 gives an explanation of how these components function together. These are colored blue in the diagram. These are introduced below alphabetically.

2.4.1 SearchGenerator

The SearchGenerator exposes a search API using the web service technology. As web services, these API return results sets in XML format and are accessible to any internet clients.

The xmdbSearch_kyoto web site relies heavily on this interface to retrieve information from the backend data-store. The data entry application also uses the web services, but interacts with the data-store via remoted objects. Client applications needing additional search functions should work with the web services rather than attacking the data-store directly.

2.4.2 DataAccessLayer

The DataAccessLayer is a set of classes that communicate with the backend database server. They talk to stored procedures implemented on the server. Much of what can be found in DataAccessLayer is very repetitive. They most follow the same implementation pattern for different data types.

2.4.3 remotingXMDBDal

The remotingXMDBDal is a very simple dotNet application which serves as the host for the remoting XMDBORLayer objects. Both Cross-MediaSearch-Generator and xmdbNG utilizes this remoting server when they access the database. Hence, when xmdbNG "retrieves" a XMDB document record from the database to edit, the document object itself is activated and executed on this remoting sever. The same applies to queries from Cross-MediaSearchGenerator.

2.4.4 SearchAPI

The SearchAPI was originally derived from nlucene (dotNet version of lucene,) an open-source full-text search engine. This also translates a lucene user input query string into corresponding SQL clause.

2.4.5 vbXMDBCustomArrayList

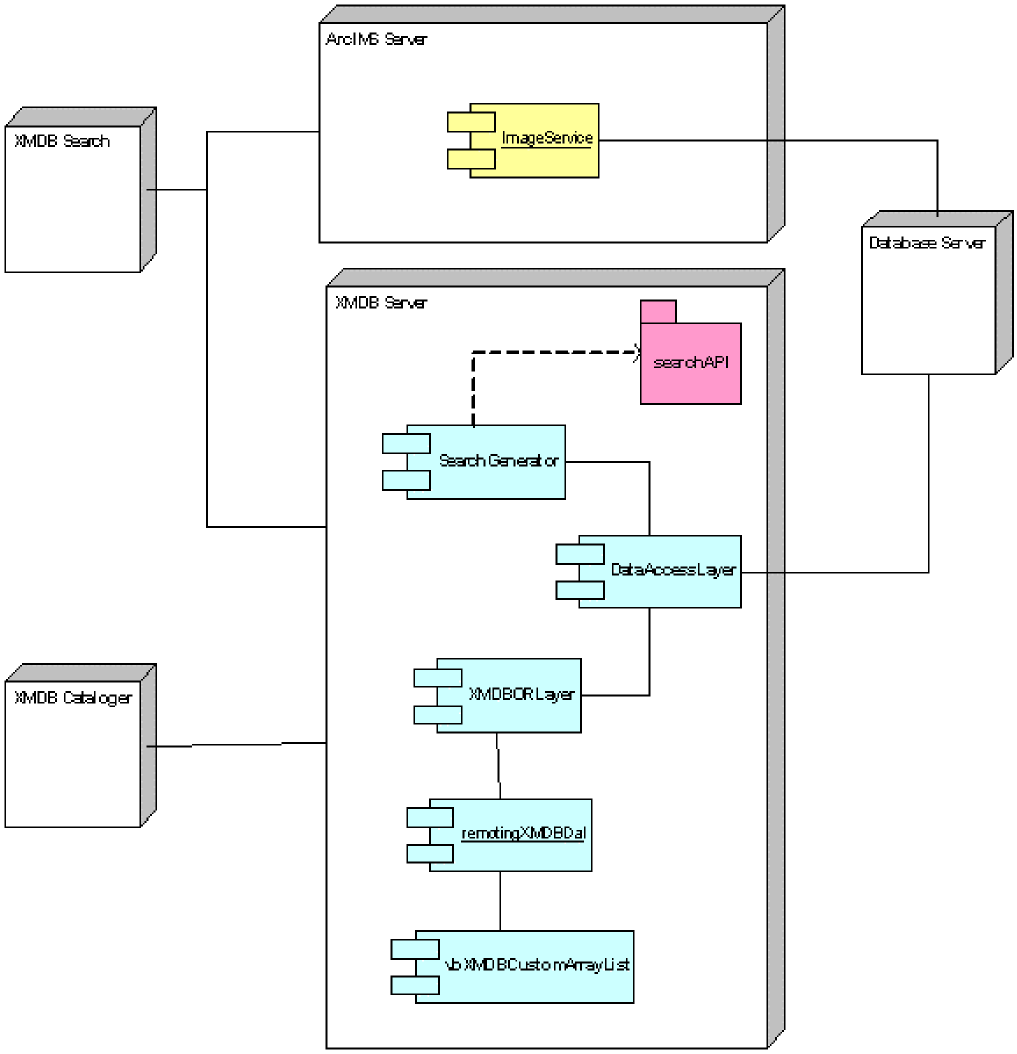

Fig. 3. Explanation of how these components function together. Functions colored blue in the diagram are introduced in sub-sections alphabetically.

The vbXMDBCustomArrayList project is a relatively small utility project to create a set of utility or typed arrays to enhance data exchange between the search web service and the web site client.

2.4.6 XMDBORLayer

This is the critical project for client application developers, since all other server components mostly just expose or retrieve data objects defined here. This project builds all XMDB resource type data elements. Due to the nature of the XMDB system, these objects has little business logic other than knowing how to interact with the persistent, that is data access, layer. For example, a client application can add “Date,” “Keyword” or “Relationship” to these objects by simply call AddDate( ), AddKeyword( ), or AddRelationship( ) respectively. These objects are responsible for calling the proper dataaccesslayer methods to store, update, or remove.

2.5 Resources

XMDB is made up of database entities called resources. A resource is simply an object, or collection of objects, catalogued and/or stored in the XMDB database. Resources have common, as well as shared, attributes or characteristics. Generic library meta-data are characteristics common to all resources in XMDB. Similarly workflow and maintenance record, access control, and usage history are all common meta-data applicable to all resources.

Different types of resources in XMDB also carry attributes specific to their individual types. For example, catalogue entries for photographic or audio records of the Niigata Cheutsu earthquake include very different descriptive meta-data than those for seismic data. While a photograph might have corresponding information on the optic or camera, an audio recording has recording speed, and seismic data have a sensors network as one of its descriptive characteristics.

Table 1. Brief description for 12 resource types using in the XMDB system.

Audio	Sound recording in and on all types of media, physical or digital. Do not include sound recordings that are on, or accompanying, motion pictures, videos or film strips.
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.
Document	A physical or digital entity containing all or a portion of a work or several works. An original or official paper relied on as the basis, proof, or support of something. A writing conveying information. Also, any paper form that has been filled in, especially one supplied by a government agency or which has legal significance.
Event	An occasion or activity, often pre-scheduled or planned; something that takes place.
Geospatial	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). Geospatial data has both spatial and thematic components.
Image	May include pictures, charts, sculpture, technical drawings, or other non-projected graphics, and/or projected graphics such as filmstrips and slides and their accompanying sound characteristics. Also use for remote sensing images defined as an image produced by a recording device that is not in physical or intimate contact with the object under study This may be a map or other image that is obtained through various remote-sensing devices such as cameras, computers, lasers, radio frequency receivers, radar systems, sonar, seismographs, gravimeters, magnetometers, and scintillation counters. Not used for documents, or images of documents, such as microforms or photocopies; those are represented by “documents” instead.
Internet Resource	A resource located on the Internet. Use ‘internet’ to catalog information about the Internet site and catalog elsewhere, as appropriate, any resources obtained from the site.
Model	Computer modeling software whose primary function is to model a certain class of physical systems. May include pre- and post-processing components and other necessary ancillary programs.
Organization	Human beings making up a group or assembly, or linked by a common interest or administrative/functional structure.
Person	Human, individual
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.
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.

There are altogether 12 resources types in XMDB. They are briefly described in the Table 1. The types are not meant to be exhaustive. Additional resource types can be introduced to the system when warranted.

Modifications have been introduced to the DRS version of XMDB to support both Japanese and English names for persons, for example. Further, Japanese names might have two different spellings, one in Kana and another in Kanji. Two tables are added to support multi-lingual names for persons. These tables are PersonNameAlias, and PersonNameLanguage_Master. PersonName-Language_Master is a valid values table to control available aliases; presently there are (1) Kana, (2) Kanji and (3) English.

2.6 Relationships

Besides the conventional thematic, or domain specific, hierarchical classification of holdings in a scientific library, XMDB implements two other: geographic and temporal. These are materialized in the spatial and timeline browsers. All resources in XMDB have some geographic and temporal characteristics. These three are examples of relationships in XMDB. But “relationship” in XMDB goes beyond these. One such relationship might connect journal papers to their authors. Another connects a dataset to a topically related web page.In fact XMDB is based on the principle that resources have relationships to other resources Relationships can be any connection between a resource being cataloged and any of the other resources in XMDB. Hence the name Cross-Media, rather than simply multi-media, database to emphasize this form of knowledge representation.

All libraries, and particularly scientific libraries, employ some sort of categorization or classification of their holdings to allow reasonable access and management. Being a multi-disciplinary repository and a multi-media database, XMDB has a very challenging “classification” task. The relationship model becomes a necessary design feature so that the body of materials on disaster management can be shared and jointly utilized by geologists, engineers, and architects, as well as, social scientists, public policy makers and historians. Although we have not quite achieved the desired set of ontologies for disaster management research, XMDB closely follows current research in the use of ontology for knowledge sharing. Ontology is defined as, “A body of formally represented knowledge is based on a conceptualization: the objects, concepts, and other entities that are assumed to exist in some area of interest and the relationships that hold among them (Genesereth & Nilsson, 1987). A conceptualization is an abstract, simplified view of the world that we wish to represent for some purpose. Every knowledge base, knowledge-based system, or knowledge-level agent is committed to some conceptualization, explicitly or implicitly. An ontology is an explicit specification of a conceptualization” by Gruber (1993).

3 XMDB Prototype

Fig. 4 shows Search and Query Tool with which a user searches a resource. As shown in Fig. 4 besides the fundamental search by the keyword input, the search from a resource type is also possible. The searched contents (the contents of description of metadata) of the resource are shown in Fig. 4. The information on the explicit relationship between resources besides the contents of description of a metadata element can be acquired, and a related resource can be derivatively followed. Moreover, as shown in Fig. 5, some raw data in a database can be displayed using installed application software. The map view of a resource is shown in Fig. 6. By using the function of GIS, a resource can be plotted in map space based on a footprint. Resources can also be displayed in time series or in calendar form (Fig. 7).

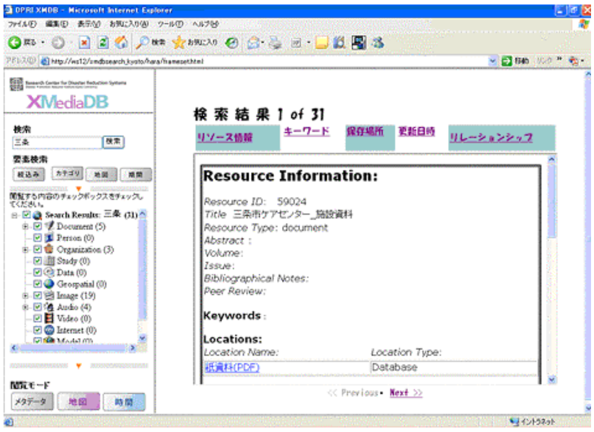

Fig. 4. Snapshot of basic search window. The upper portion in the left flame is input field for search keywords. The lower portion in the left flame is selection window for resource types (optional). The

4 Concluding Remarks

In this research, the Cross-Media database as RSII (Reserch Support Information Infrastructure) was developed for responding to the needs of a man of business and a researcher who are engaged in disaster prevention. The feature should be a database for a multi-resource type. In order to unify a multi-resource type resource, be based on the metadata set of

Fig. 5. Snapshot of the view flame for raw data. In this case PDF file is displayed using Adobe Acrobat Reader.

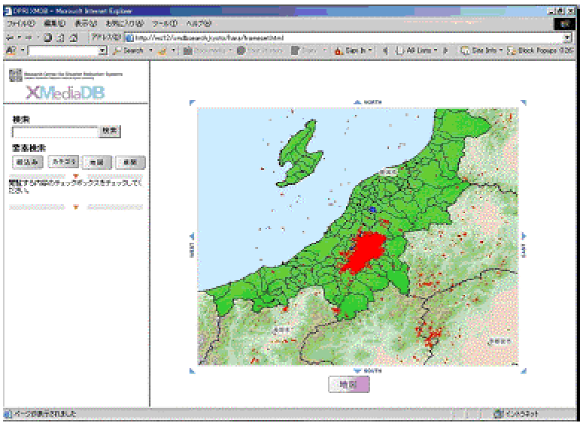

Fig. 6. Snapshot of the map view of a resource. Dots represents the epicenters of aftershocks for 2004 Niigata Chuuetsu Earthquake.

international standards. The explicit relevance between resources (relationship) is storable. It is the information retrieval based on time, space, and a theme being possible, carrying the new display function of the search result by time and space, etc. The system design was performed while performing the concept design based on these. In the system design, the architecture of XMDB was designed from four-layer structure, and the database schema about metadata was designed simultaneously. Finally the functions of XMDB were offered via user interface. Moreover, the actual proof examination is started through the trial using the data of various resource types about heavy rain disaster in Niigata Prefecture in 2004 and 2004 Niigata-ken Chuuetsu Earthquake. The left-behind subject is the necessity of customizing about the keyword (this being called controlled vocabulary). The present XMDB follows the global standard about the data format. Specified data formats for disaster science, such as SAC and WIN for seismogram format,

Fig. 7. Snapshot of time series (top) and a calendar form (bottom).

should be considered. The XMDB proposed in this paper is a research infrastrucutre for a man of business and a researcher on disaster prevention. The structure of a database, the raw data and metadata stored in a database grows and develops by positive participation and use of the disaster prevention researcher which participates as a stakeholder. Not management of the data/information depending on an old researcher individual but systematic structure that shares the measure and knowledge of a researcher should raise the quality of disaster prevention research.

Acknowledgements

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

Genesereth, M. R., & Nilsson, N. J. (1987): Logical

Foundations of Artificial Intelligence. San Mateo, CA: Morgan Kaufmann Publishers.
Gruber, T. R. (1993): A translation approach to portable ontologies. Knowledge Acquisition, 5(2):199-220.
Kawakata, H., Kawata, Y., Hayashi, H., Tanaka, T., Topping, K.C., Yamori, K., Yoshitomi, P., Urakawa, G., and Kugai, T. (2004): Building an Integrated Database Management System of Information on Disaster Hazard, Risk, and Recovery Process. Annuals of Disaster Prevention Research Institute, Kyoto University, 47(C), 89-92.

災害ハザード・リスク・復興過程等に関する情報の統合型データベース・システム（クロスメディア・データベース）の構築（２）

川方裕則・ポール吉富・浦川豪・Kelly CHAN・松浦秀起・
辰己賢一・原武士・阿草宗成・林春男・河田惠昭

要旨

災害に関する適切なトータル・リスク・マネージメントには、災害の予知・予防・情報伝達・災害履歴・災害復興過程などの様々な情報、データの共有が不可欠である．これら情報、データを有機的に広く活用できるようにするために、本研究では防災に関する統合型データベース・システム（クロスメディア・データベース）の構築が進められており、2004年に新潟県を中心に発生した豪雨災害、中越地震災害を例としたプロトタイプが完成した．

キーワード: データベース，防災，クロスメディア・データベース，メタデータ，リレーションシップ・研究支援情報基盤