Development of a Science Database System Applicable to Various Access Restrictions

DEVELOPMENT OF A SCIENCE DATABASE SYSTEM APPLICABLE TO VARIOUS ACCESS RESTRICTIONS

Y Takata*, Y Kasahara, and T Matsuhira

Kanazawa University, Kakuma-machi, Kanazawa-shi, Ishikawa, 920-1192 Japan
**E-mail:* yosihiro@kenroku.kanazawa-u.ac.jp

ABSTRACT

We developed a general-purpose database (DB) system that manages and opens experimental and/or observational data accumulated in universities and academic institutes. This platform works as a web-DB management system in which databases can be easily managed without special skills and facilities. By defining a group manager and data manager, the proposed system defines a flexible access restriction for each user and each unit of datasets under the control of these managers. We demonstrate how a variety of web-DBs are appropriately integrated under one management system in such situations in which each web-DB has a different user interface in its search and data distribution functions and is designed with a different language (or script) and connection method to its DB. In spite of the diversity of web-DBs, the proposed system is highly suitable for practical use.

Keywords: Web-DB management system, Access policy, Access restriction, Authentication, Authorization, Experimental and observational data, Global environment

1 INTRODUCTION

As the information society develops, universities and academic institutes are required not only to accumulate valuable academic materials but also to open them to the public. Original research materials such as experimental and observational data are very important. However, the accessibility of these data is not always easy even though there are many requests for the data from scientists inside and outside the institutes.

In academic fields such as meteorology, oceanology, seismology, and solar-terrestrial science, it is quite important to cross-refer and make comprehensive analyses using various kinds of data obtained by many observatories and institutes in order to clarify the global picture of the environment. With the evolution of computer and network engineering, it became mandatory to use multiple kinds of data for further understanding the global environment beyond the boundaries of research fields. In addition, ordinary people also wish to refer to these data via the Internet.

In recent years, many countries have begun to construct global database systems for integrated data management and to open the data to the public. For example, the National Space Science Data Center (NSSDC) in the National Aeronautics and Space Administration (NASA) is responsible for constructing systems for accumulation, management, and release of data about the U. S. space science mission. The Data Archives and Transmission System (DARTS) in the Japan Aerospace Exploration Agency (JAXA) manages a unified database system for space science in Japan. However, these systems generally manage the data obtained under national projects, and the data access policy is basically uniform for each project.

On the other hand, in our study, we focus on datasets under the control of small research groups, such as laboratories in universities. Portions of these datasets have already been opened to the public through their original database systems, but it is difficult to unify them because of the variety of user interfaces. In addition, the majority of these datasets are still not efficiently accessed because the data managers are not necessarily experts in database system management. It is also noted that data owners and their research groups have a priority to use their data exclusively for a certain period. However, they are also responsible for releasing the data after that period has elapsed or their research has progressed to a certain point. In order to release these datasets easily, it is necessary for the database system to be able to control data access according to a formulated policy.

With the above as background, we have developed a general-purpose database system that manages and opens experimental and/or observational data accumulated in universities and academic institutes. As has been

mentioned previously, the most critical point in the system development is flexibility of data access control. There are a variety of data policies: fully open access, partially open (for example, opening data more than 3 years old), open low resolution data only, and/or subscribers only. In general, these policies depend on the particular datasets, institutes, laboratories, and/or academic fields.

We have developed a common platform system for a "web-DB management system" in which databases are easily managed without special skills and facilities. In order to realize such a common platform, our system is designed based on an open web-DB system (web-DB). The web-DB management system defines flexible access restrictions for each user and each unit of datasets under the control of the data manager and group manager. The advantage of our platform is that it is possible to manage both newly constructed web-DBs and existing web-DBs by putting them under the control of the managing system and to reduce the burden of the data manager in maintenance of his database. It is also noted that data policy can be given separately to each unit of datasets and/or users in a unified system, in which multiple laboratories, academic fields, and DBs managed by different data managers are inter-mixed.

In this paper, we introduce our web-DB management system and demonstrate an application example using two kinds of experimental and observational datasets: 1. datasets containing plasma and radio wave measurements from the Akebono satellite and their ancillary data and 2. datasets containing gravity anomalies in the Japan Islands. In sections 2 and 3 respectively, we introduce a design concept and specifications of our web-DB management system. A demonstration of the web-DB management system is given in section 4. Finally, we conclude our study in section 5.

2 DESIGN CONCEPT OF THE DEVELOPED WEB-DB PLATFORM

2.1 Authentication and authorization

In our system design, every user is authenticated with a user ID and password and authorized to obtain datasets on a web-DB. Authentication is used to identify the user and confirm his identity, and authorization gives authority or permission to access the resources available to the authenticated user. Because users have a variety of requirements for access to the contents stored in the system, it is necessary for the data manager to identify these users and control their access.

In particular, it is impossible to define data access policy as fully open access or access for subscribers only. It is not out of the ordinary for a data manager to have exclusive use of his group's data for a certain period before the data are released or to release only low resolution data. In order to meet such requirements, it is necessary to divide the data access into stages, specifying what level of data can be released, when it can be released, and to whom. The schematic concept of these stages is shown in Figure 1. Followings are examples from the present system:

- Only the data owner (observer who acquired the data) is permitted to access the raw data.
- The data owner and the collaborating research groups are permitted to access the calibrated data.
- All users are permitted to access the summary-plots and/or low-resolution data.

The above three cases might be nominal examples. We also take into account the following special case:

- A portion of users are permitted to access a specific portion of the data for special research projects, temporal projects, or campaign observations.

It is important to reduce the burden on each database manager (primarily the data owner) to the minimum by deciding what data should be released and when this should be done. In the following sections, we introduce the design concept of our platform and the relationship between the management system and web-DBs.

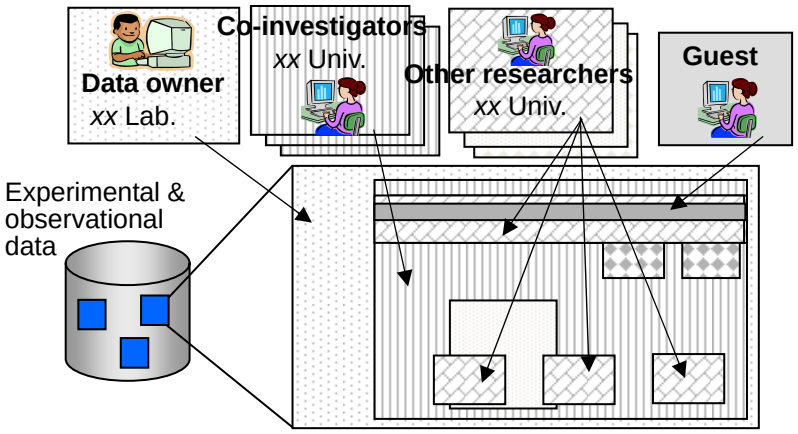


Figure 1. Various access restrictions

2.2 Policy on user management

In this section, we define a user management policy. The left panel in Figure 2 shows the hierarchical structure of users in our system. The users in the top level are data owners who are actually responsible for the experimentation or observation project. The users in the second level are co-investigators who are involved in the project. The third level users are other researchers or scientists in academic fields. The fourth and fifth user-levels are other people who are interested in the datasets. Although we show only five levels in the panel, technically, our system is able to define more levels. We assume also that every user belongs to a research group that is working on a research project. However, one user might belong to more than one group, and one research group might be involved in more than one research project. We define such situation schematically as shown in the right panel in Figure 2. In the panel, researcher "c" belongs to research groups "G1," "G2," and "G4" simultaneously, and research group "G4" is involved in research projects "B" and "C".

On the other hand, a group may use its data exclusively for a certain time period before the data are released in stages as described in Section 2.1. For example, data owned by research project "B" in Figure 2 are initially released only to users who belong to research groups "G2" and "G4." They are subsequently released to other researchers belonging to groups "G1" and "G3," which are not involved in project "B" but are in the same academic field. In some cases, the data are widely released to general users following the hierarchical structure shown in the left panel in Figure 2.

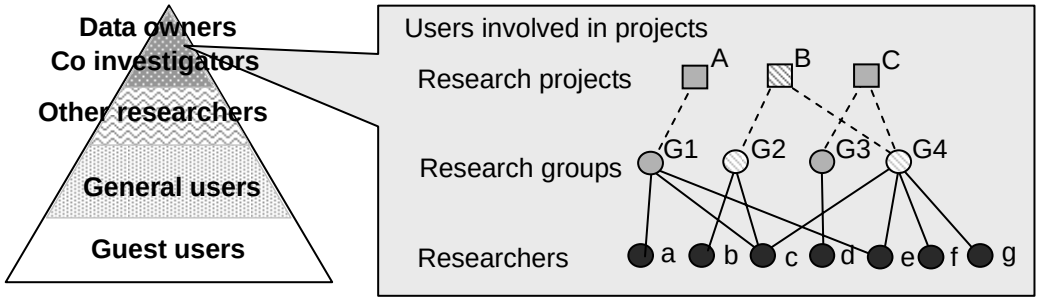


Figure 2. Concept of user management

2.3 Policy on data access control

In general, data access policy is not necessarily same for all data stored in a database. It depends on the type of data and on the individual users as described in the previous section. In our system, we assume that the following cases will be used frequently for data access control:

(a) Apply an access control policy to every unit of service implemented in a web-DB.
(b) Apply an access control policy to every unit of datasets in a web-DB. Examples: name of DB, name of DB-table, name of column in a DB-table.

(c) Apply an access control policy to every species of data. Examples: raw data, calibrated data, summary-plots, name of instrument.

(d) Apply an access control policy by the range of data such as latitude and longitude, name of area, etc.

(e) Apply an access control policy by range of dates when the experiment or observations were performed.

Items (a) and (b) are related to system design, while the other items originate from the characteristics of the datasets. It is easy to implement items (a), (b), and (c) because modifications of system design in web-DBs are not required. For items (d) and (e), however, it might sometimes be necessary to touch up the operations (such as SQL statements) for search and extraction of requested data in a web-DB.

2.4 Relationship between the web-DB management system and web-DBs

In order to meet independently all requirements for user management and data access control in the existing web-DBs, it is necessary to authenticate and authorize users under quite complicated data access control policies. This imposes a heavy burden on the data owner from both technological and operational points of view. In our system design, however, these roles are assumed by the web-DB management system, and the web-DBs are responsible only for providing their data to users authorized by the management system. The roles of the web-DB management system and the web-DBs are summarized as follows:

[Roles of web-DB management system]

- Authenticate and authorize user.
- Manage information tables necessary for authentication and authorization, such as information on users and research groups registered in the system, web-DBs integrated in the system, and conditions for authorization.
- Manage access records.
- Provide web interfaces associated with these processes.

[Role of web-DBs]

- Search and distribute experimental and/or observational data at the request of authorized users.

In order to administer the web-DB management system and web-DBs separately, we define three kinds of administrators (managers) who play different roles in system management. First of all, it is necessary to define a system manager responsible for the entire web-DB management system. This system manager controls all management issues, such as user registration and data access control. However, it is difficult for the system manager to administrate everything in a large-scale system. On the other hand, it is appropriate for each data owner to manage the data access policy of his own web-DB. Likewise, it is also appropriate for a representative of a research group to manage the members belonging to his group. The definitions of these managers are summarized as follows:

[System manager]

- Manages the entire web-DB management system
- Assigns data managers and group managers

[Data manager]

- Manages the data access control of his own web-DB through a web interface provided by the web-DB management system. The preferred data manager is the data owner or a person involved the research project team. (See sections 3.2 and 3.4.2)

[Group manager]

- Manages user registration of his research group members through a web interface provided by the web-DB management system. The preferred group manager is a representative of the laboratory or research group. (See sections 3.2 and 3.4.1)

2.5 User Interface and Process Flow

In this section, we introduce the general concept of user interface and process flow in the web-DB management system. Figure 3 shows a schematic flow when a user requests access to a web-DB system. The flow of responses by the web-DB management system is as follows:

- When a user tries to access the web-DB management system, an authentication page is displayed until his authentication is completed. The web-DB management system requires a user ID and password to complete authentication (Figure 3(a)).
- After authentication, a selection page of web-DBs is displayed. When the user selects one of the web-DBs listed on the page, the web-DB management system compiles an accessible data list from the user information and the data access policy of the selected web-DB (Authorization) (Figure 3(b)).
- The user is redirected to the selected web-DB, and a list of data accessible to him is given to the web-DB. Based upon the list, the web-DB identifies the accessible data and sends them to the user (Figure 3(c)).
- The user is permitted to access the data under the data access control of the web-DB management system (Figure 3(d)).
- The user is required to repeat steps (b) to (d) to access other web-DBs.

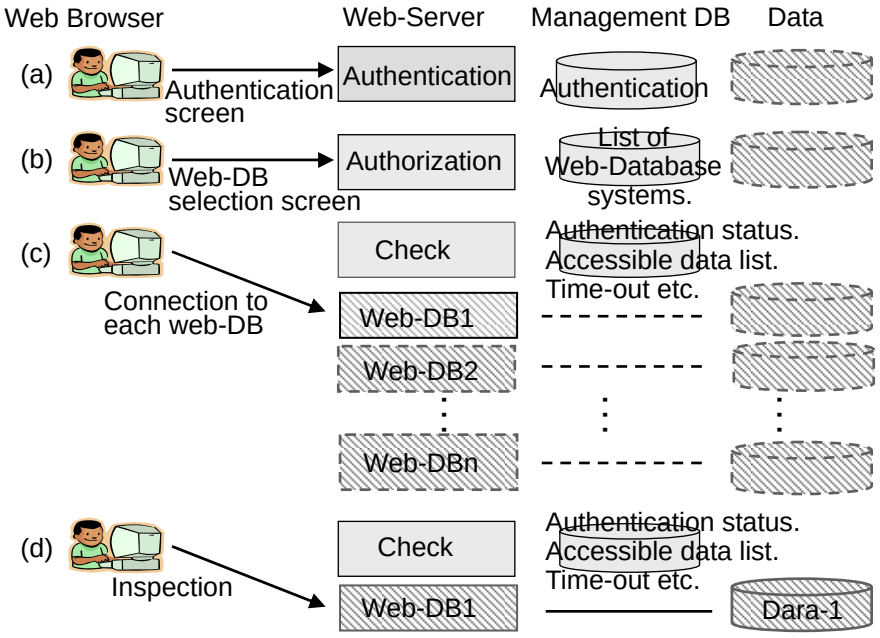


Figure 3. Summary of system operations

3 Specifications

3.1 System configuration

Figure 4 is a summary of the system. This system is designed to separate a Web server from DB servers. The user interfaces are roughly divided into functions for managers (system, data, and group) and ordinary users. Furthermore, functions for managers are divided into user management functions for system and group managers and data management functions for system and data managers.

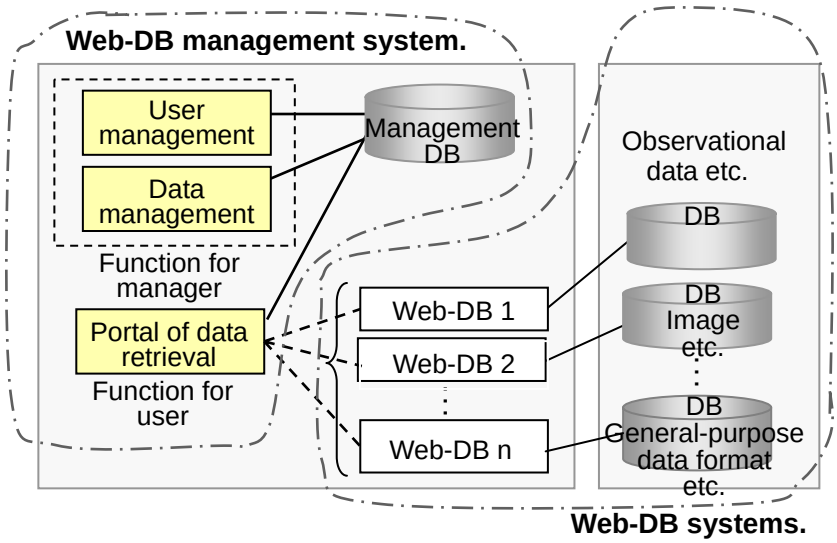


Figure 4. Summary of the system

3.2 DB specifications for management functions

In our web-DB management system, we utilize a relational database system. An entity relationship diagram is shown in Figure 5. The tables are roughly divided into two groups according to their purpose: (1) for user management and authentication and (2) for authorization.

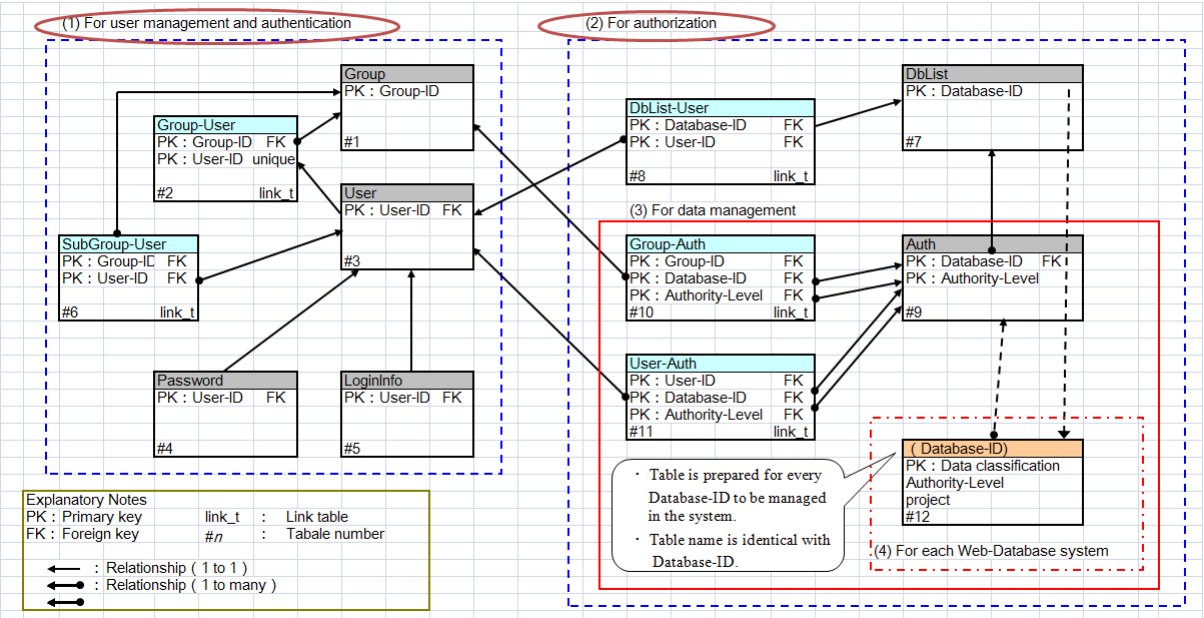


Figure 5. An entity relationship diagram of the Web-DB management system

The tables used for user management and authentication are: #1, a table of group profiles (Group); #3, a table of user profiles (User); #4, a table of passwords for user authentication (Password); #5, a table for the management of authentication information (Longininfo); and #2 and #6, tables to associate the users with groups and subgroups (Group-User and Subgroup-User), respectively. As stated in section 2.2, the system is able to register one user with multiple groups (subgroups). Attributes of users are registered in the User table (#3), which is used for the basic authority of the user. Attributes of group managers are also included in table #3, and thus, the user who is assigned to be a group manager is able to manage his group.

The tables used for authorization are: #7, a table to manage web-DB information (DbList); #8, a table to assign a

data manager to each web-DB (DbList-User); #9, a table to manage each user or group's data access policy (Auth); #10 and #11, tables to describe access authorities of groups and users, respectively (Group-Auth and User-Auth); and #12, a table to manage the information delivered to each web-DB (Database-ID). #12 is separately defined for each web-DB.

Data managers of each web-DB assigned in DbList-User (#8) can manage the authority of data access control to his web-DB on behalf of a system manager. Auth (#9) also controls the authority for special research projects or temporal research projects (See section 2.1) as well as basic data access policy described above.

3.3 Authorization module for web-DBs

We provide a basic authorization module for web-DBs to be integrated into the management system. This module is imported into the web-DB and can be customized according to the environment of each web-DB.

Although various languages (or scripts) are used in packaging the basic authorization module for each web-DB, we include a module for PHP and class modules for Java-based languages (JSP, Servlet) as they are used so frequently.

In order to integrate a web-DB into a web-DB management system, it is necessary to install the basic authorization module into the web-DB and to separate the web-DB functions into Web interface and DB server. In the separation process, we take into account the following two cases: (1) a central data center, managing a web-DB management system, is responsible for serving as a Web interface, and the research group owning the original data manages the DB server of its datasets and (2) a central data center manages both Web and DB functions as a proxy for the research group. Case (1) is recommended because the datasets are managed by the data owner in the research group who is familiar with the content of the datasets and is responsible for registration and upgrading of the datasets. Case (2) is applied if a research group has difficulty managing its datasets alone.

3.4 Management functions

3.4.1 User management functions

Table 1 shows a list of user management functions: management of group information (1–3), management of user information (4–6), management of subgroups (7), and access log management (8). As was mentioned in section 3.2, a user's basic authority is contained in his user attributes. We define four categories of basic authority: system manager, group manager, general user, and guest user. A system manager can assign any user to be a group manager. A user assigned as group manager can manage user information in his group (excluding subgroups) on behalf of the system manager.

Table 1. List of functions for user management

	Function	Roles of system manager	Roles of group manager
1	Modification of group info.	∨	∨
2	Deletion of group	∨	
3	Creation of group	∨	
4	Modification of user info.	∨	∨*
5	Deletion of user	∨	∨*
6	Creation of user	∨	∨*
7	Sub group	∨	∨
8	System log management	∨	

* It is prohibited to create, delete, or modify the user registration information of system manager or group manager.

3.4.2 Data management functions

Table 2 shows a list of data management functions. DB master (1) registers the attributes of the web-DBs in table #7 (DbList) such as the name of the web-DB, its explanation, and its URL.

Data management (2) assigns a web-DB data manager. Using this function, the data manager of each web-DB is registered in table #8 (DbList-User), and the user assigned as a data manager is entrusted with managing the access authority of his web-DB (See section 2.4). Both system manager and data manager are permitted to use functions from 3 to 7 in Table 2 in order to manage data access.

Authority master (3) manages table #9 (Auth), which defines groups' and users' basic access. This function is also used to define the special authority (project authority) defined in table #9 (Auth). Table 3 shows an example of the list of basic access authorities given by default when a web-DB is first put under the control of a web-DB management system. As shown in Table 3, authority levels 01–04 and 09 are used for basic access authority by default, but the definition of the authority levels can be customized for every web-DB. In the case of the definition of special authority (project authority), authority levels from AA through ZZ are used so that 676 types of special authorities may be defined for each web-DB.

Group authority (4) and user authority (5) are used for giving data access authority to groups and users, respectively. Group and user definitions are registered in table #10 (Group-Auth) and #11 (User-Auth), respectively. The definition of User-Auth takes precedence over Group-Auth. For a user whose data access authority is not defined in table #11 (User-Auth), data access authority is given by the group authority #10 (Group-Auth) of his research group. For a user who is not defined by user authority or group authority, a default access authority (currently defined as authority level 04) is adopted.

Individual DB management functions are used by a unit of datasets to apply data access control in a web-DB. Table 4 shows an example of individual DB management. In the table, "obsXXXX" is a unit of datasets that can independently define data access policies. In this case, for example, the dataset "obs1989" is accessible by all users including guest users, while the dataset "obs1990" is accessible only by users whose access level is higher than level 04 (other than guest users) or users who have special authority (project authority) from AK. Likewise, "obs1991" is accessible only by users whose access level is higher than level 02 (the owner and collaborating researchers) or authorities of AK or CE.

DB list survey (7) browses information on the web-DB necessary for management.

Table 2. List of functions for data management

	Function	Roles of system manager	Roles of data manager
1	DB master	↓	
2	Data management	↓	
3	Authority master	↓	↓
4	Group authority	↓	↓
5	User authority	↓	↓
6	Individual DB management	↓	↓
7	DB list survey	↓	↓

Table 3. Basic access authorities

Authority level	Name	Available data	Meaning
01	Basic level 01	All data	System manager
02	Basic level 02	Part of data	Data manager and co-investigator
03	Basic level 03	Part of data	Collaborator
04	Basic level 04	Part of data	General user (Default)
09	Basic level 09	Part of data	Guest user

Table 4. An example of individual DB management

Unit of datasets	Authority level	Project1	Project2	Project3
obs1989	09	AK		
obs1990	04	AK		
obs1991	02	AK	CE	
obs1992	02	AK		
obs1993	02	AK		
obs1993	02	AK		
obs1994	02	AK		
obs1995	02	AK		
obs1996	02	AK		
obs1997	02	AK		
obs1998	02	AK		
obs1999	02	AK		
obs2000	02	AK		
obs2001	02	AK		
obs2002	02	AK	CE	

4 Demonstration

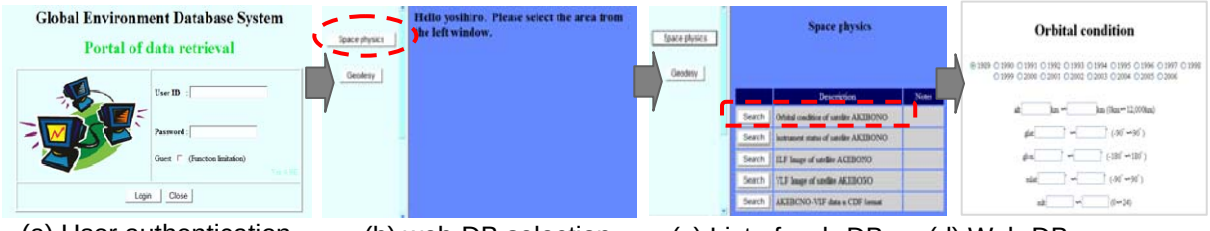
4.1 Development environment

In this section, we demonstrate an application of the proposed web-DB management system constructed for the "Global Environment Database System," in which digital data on global environment measurements and observations accumulated in our university. The configurations of the developed system are described as follows:

- Hardware: We adopted a rack-mountable server that implemented a CPU of Intel Xeon with 3.40 GHz and RAM of 2 GB.
- Software: We adopted a Linux operation system (OS) in which Apache and PostgreSQL were installed as web server and DBMS, respectively. The PHP language was used for management system development.

4.2 Integration of web-DB into the web-DB management system

We constructed a web-DB management system and integrated several new and existing web-DBs in our university for demonstration. Figure 6 shows snapshots of web interface pages of the developed management system. Each snapshot corresponds to the process shown in Figure 3.



(a) User authentication (b) web-DB selection (c) List of web-DB system (d) Web-DB: Orbital Condition of the Akebono satellite

Figure 6. Snapshots of the developed web-DB management system (https://wwwdb01.db.kanazawa-u.ac.jp/DB-E/pub/db/)

In the demonstration, we integrated six kinds of web-DBs in two academic fields, space physics and geodesy, under the control of the web-DB management system. A list of the web-DBs is shown in Table 5. In order to demonstrate how varieties of web-DB are appropriately integrated under one management system, we purposely made a situation where each web-DB has a different user interface in its search and data distribution functions and is designed with a different language (or script) and connection method to its DB. Detailed differences in user interface are described as follows:

- Table 5: 1, 2, and 6 - Search results are given directly by responding to a query to the DB in which datasets are stored.
- Table 5: 3 and 4 - A meta-DB that manages meta-data of image datasets (.png) is searched first, and images corresponding to search results are displayed. In the process, a requested image dataset is transferred to the Web server via the data server using rsync over ssh.
- Table 5: 5 - A meta-DB is searched first. The corresponding datasets are described in Common Data Format (CDF), and the search results are shown through an additional process of drawing, displaying, and downloading the corresponding CDF files onto the Web server. We adopted XML/Web Service (SOAP over http) for communication between DB and Web servers.

The integration of each web-DB under the control of the web-DB management system was performed as follows:

a. Define a policy of access restriction to each web-DB (See section 2.3).
b. Construct/reconstruct the web-DB.
   - In case of an existing web-DB: reconstruct the web-DB in advance if it is necessary to optimize DB-tables and processing script.
   - In case of a new web-DB: design and package the system according to the access restriction policy.
c. Customize the authorization module
d. Input information on access restrictions (according to the defined access policy) from the management function of the web-DB management system.
e. Operation test

Table 5. Summary of each web-DB

	DB-ID	Contents	Field	Lang./ DB	Method of connection	Access restrictions	Volume of data
1	ake-obt	Orbital condition of the Akebono satellite	Space physics	PHP PostgreSQL	Port forwarding by ssh	open/close/conditional： access control is given by a unit of observation year	~ 18 million records
2	ake-obs	Instrument status of the Akebono satellite	Space physics	PHP PostgreSQL	Port forwarding by ssh	open/close/conditional : access control is given by a unit of observation year	~ 1000 million records
3	ake_elf_image	ELF spectrogram measured by the Akebono satellite	Space physics	PHP PostgreSQL	Port forwarding by ssh , rsync over ssh	open/close	~ 120,000 image files
4	ake_vlf_image	VLF spectrogram measured by the Akebono satellite	Space physics	PHP PostgreSQL	Port forwarding by ssh , rsync over ssh	open/close	~ 120,000 image files
5	ake_cdf	VLF data measured by the Akebono satellite in CDF format	Space physics	JSP, Java PostgreSQL	XML/Web service(SORP over http)	open/close/conditional : access control is given by sub-instrument	~ 24,000 CDF files
6	Gravity	Gravity anomaly of the Japanese Islands	Geodesy	PHP PostgreSQL	PostgreSQL :5432port	open/close/restrictions on data download	~ 550,000 records

4.3 An example of user interface

Figures 7 and 8 are examples of snapshots of the web interface of a web-DB after authentication by the web-DB

management system.

Figure 7 contains snapshots of three types of query interface of the instrument status database of the Akebono satellite (DB-ID #2 in Table 5). We defined a data access policy by a unit of an observation year on this web-DB (See Table 4). Under the control of the web-DB management system, units of datasets are accessible by a user who has basic or special project authority. Panel (1) illustrates when a data owner or a co-investigator accesses the web-DB (basic level 02 in Table 3). Panel (2) is a web interface for a user who is given special authority (project authority), and panel (3) is that of a guest user (basic level 09 in Table 3).

Figure 8 contains snapshots of a DB selection page for the Akebono VLF database (DB-ID #5 in Table 5). We defined a data access policy by a unit of a subsystem of instruments for this web-DB. Panel (1) illustrates a data owner or a collaborating researcher accessing the web-DB (basic level 02 in Table 3). The data division displayed uses a controlled method similar to that of Figure 7. Panel (2) is the web interface for a cooperating researcher (basic level 03), and panel (3) is that for a general user (basic level 04). This web-DB does not give access authority to a guest user (panel (4)).

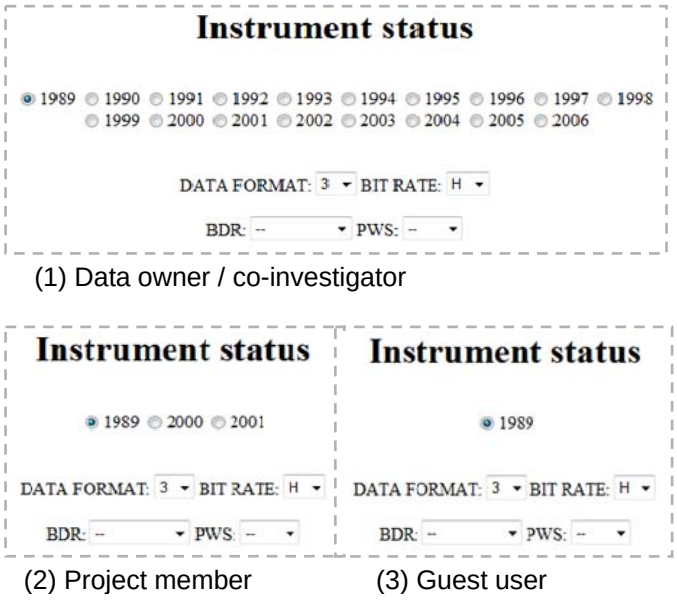


Figure 7. Restriction examples by a unit of observation year

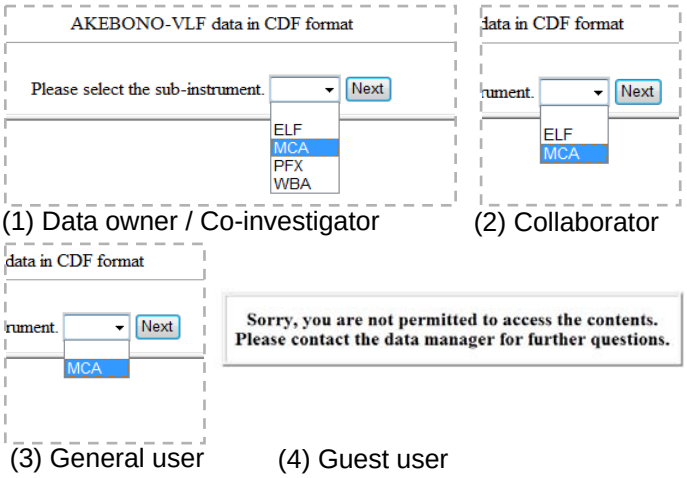


Figure 8. Restriction examples by a unit of observation equipment

5 CONCLUSION

In the present study, we developed a common platform system for a "web-DB management system" in which databases are easily managed without special skills and facilities. It is important to note that our system defines flexible access restriction for each user and each unit of datasets under the control of a data manager. It is also possible to integrate both newly constructed web-DBs and existing web-DBs by putting them under the control of the managing system, which reduces the data manager's burden. Data policies are given separately to each unit of datasets and/or users in a unified system, in which multiple laboratories, academic fields, and DBs managed by different data managers are inter-mixed.

In the demonstration, we integrated six kinds of web-DBs in two academic fields, space physics and geodesy, under the control of the web-DB management system. In order to demonstrate how varieties of web-DB are appropriately integrated under one management system, we purposely constructed a situation in which each web-DB had a different user interface for its search and data distribution functions and was designed with a different language (or script) and connection method to its DB. In spite of the diversity of web-DBs, the proposed system is highly suitable for practical use and has achieved its initial goal of reducing the burden of the web-DB manager. In addition, it is highly applicable to web-DBs in a variety of academic fields because it has versatile specifications that do not depend on the characteristics of datasets and the implementation method of a web-DB.

6 ACKNOWLEDGEMENTS

This research was supported by the Grant-in-Aid for Scientific Research, Japan Society for the Promotion of Science (20510006).

7 REFERENCES

National Space Science Data Center (NSSDC), NASA. Retrieved from the World Wide Web on Dec 19, 2008: http://nssdc.gsfc.nasa.gov/

Data Archives and Transmission System (DARTS), JAXA. Retrieved from the World Wide Web on Dec 19, 2008: http://www.darts.isas.jaxa.jp/index.html.en

Common Data Format (CDF), Goddard Space Flight Center of NASA. Retrieved from the World Wide Web on Dec 19, 2008: http://cdf.gsfc.nasa.gov/

Global Environment Database System, Kanazawa University. Retrieved from the World Wide Web on Dec. 30, 2009: https://wwwdb01.db.kanazawa-u.ac.jp/DB-E/pub/db/

(Article history: Received 7 April 2009, Accepted 2 January 2010, Available online 13 January 2010)