International Seabed Authority ISBA/9/LTC/3

Legal and Technical Commission

Distr.: Limited
29 April 2003

Original: English

Ninth session
Kingston, Jamaica
28 July-8 August 2003

Status of the International Seabed Authority's central data repository on marine mineral resources

Prepared by the Secretariat

1. At its previous session, in July 2002, the Legal and Technical Commission noted the importance of the establishment of a central data repository as a core activity of the Secretariat and requested the Secretariat to provide it in 2003 with a report on the status of the database and an inventory of data holdings. The present document is submitted in response to that request.

I. Marine minerals database

2. Data and information on marine mineral resources, while known to exist, are dispersed within various organizations and companies worldwide, in various formats and standards, and are usually not readily accessible to potential users.

3. In 2000, the Secretariat of the Authority decided to improve this situation by establishing a central data repository (CDR). The objective of the CDR is to collect and centralize all public and private data and information on marine mineral resources. This should allow for the creation of uniform data formats and useful summaries in a central location, readily accessible to users. The Authority intends that the CDR should:

(a) Be accessible to all members of the Authority via the Internet;

(b) Display the acquired data and information and make possible the elaboration of listings, graphs and maps;

(c) Include quantitative mineral assessments;

(d) Enable the Authority to process information for the purposes of preparing technical reports, producing data on CD-ROM and inputting data on the web site.

4. As a preliminary phase, the Secretariat has assembled information concerning the form and availability of relevant data within 18 institutions worldwide. It

03-34107 (E) 020603
0334107

commenced the project in 2001 with the collection of data and information related to polymetallic nodules and ferromanganese crusts. The data collection was enriched in late 2002 with the receipt of data pertaining to hydrothermal vent systems and polymetallic sulphides. The data sets have been structured and integrated into the CDR and are accessible via the Internet at www.cdr.isa.org.jm or through a link on the Authority's web site at www.isa.org.jm.

5. The status of the CDR as of April 2003 is described in the following paragraphs. A simplified architecture of the data repository is presented in the annex to the present document.

A. Ferromanganese crusts

6. The Secretariat has collected all available data sets from two main sources within the United States Geological Survey (USGS): the Survey Headquarters at Reston, Virginia, and the USGS Office at Menlo Park, California. The ferromanganese data are grouped into the following sets:

(a) A geochemical data set containing location, depth, crust thickness and all geochemical data (over 70 elements), including the analysis methods and the major elements, for a total of 3,533 records;

(b) A sample data set containing ancillary data which describe the source of the data and the characteristics of the samples (3,533 records);

(c) A reduced data set containing single entries for each location represented in the data set. The original data files include multiple analyses for single sample collections and sometimes single samples. This set averages together the data for all such replicate data (1,225 records);

(d) A major elements set which is a subset of the geochemical data set, but restricted to the nine major elements accounting for 90 per cent of the materials in the nodules (Al, Co, Cu, Fe, Mn, Ni, Pb, Si, Zn). This set contains 3,533 records.

B. Polymetallic nodules

7. The Secretariat collected polymetallic nodule data available within the United States Government National Geophysical Data Center (NGDC). The data set consists of primary data files, ancillary data files and information on the sources of the data. The data are organized in sets similar to those for the ferromanganese crusts except that there is no reduced data set.

(a) The geochemical data set contains location, depth and all geochemical data (over 60 elements), including the major elements, with their respective analysis methods. It currently contains 2,753 records;

(b) The sample data set contains ancillary data available for these samples (2,753 records);

(c) The major elements set includes the nine elements that account for more than 90 per cent of the materials in most seabed polymetallic nodule deposit nodules (Al, Co, Cu, Fe, Mn, Ni, Pb, Si, Zn).

8. Other data sets provide information on the sources of the data in these three primary sets. This includes a cruise data set, containing information pertaining to 186 polymetallic nodule exploration-related cruises.

9. A total of 2,753 subsample records have been selected from the 5,662 subsamples of the original NGDC database. The rationales for exclusion and inclusion are discussed in section 4 of the online documentation.

C. Analytical procedures and elementary statistics

10. The original data sets are composed of data produced by several scientists using a variety of analytical techniques, including various wet chemical and spectroscopic methods with appropriate standardization and precision checks. Initial sample preparation procedures and analytical procedures are fully presented in the online documentation and an analysis method column is linked to the specific analytic procedure used. Because different researchers reported their results for the same variable using different units, any particular variable can be represented in the data file as weight per cent (%), parts per million (ppm), parts per billion (ppb) or parts per trillion (ppt). In fact, many of the variables in the original data file are represented in several different units. For example, nickel (Ni) is reported alternately in %, ppm and ppb.

11. To permit easy mapping and other manipulation of the data, all values of each variable have been converted to a common unit basis. Units chosen for each variable were selected for convenience, to represent the values as relatively small numbers larger than 1, but limited to %, ppm and ppb. For units other than %, the variable name, in the header record of the CDR files, includes the appropriate abbreviation to denote the units used for the variable (e.g. Al pc, As ppm, Au ppb).

12. Statistics and histograms derived from the geochemical data set are equally available in the online documentation. The geographical distribution of the data is shown in figure 4-2 in that documentation. These data summaries permit first-order examination of the key data variables and show the number and distribution of the available data. They also provide a rough quality control tool that is useful in the identification of obviously incorrect data. All these data sets are linked by a single identifier code for each subsample analysed.

D. Hydrothermal vent systems and sulphide database

13. In late 2002, the Secretariat acquired from the Geological Survey of Canada a validated set of data on the worldwide distribution of seafloor polymetallic sulphides. During the first quarter of 2003, the Secretariat integrated that data set into the CDR. The related data are grouped into four functional sets:

(a) The primary set (geochemical data) contains geochemical analyses for 2,640 samples of seafloor polymetallic sulphides and related hydrothermal precipitates from 69 different sites worldwide. The actual compilation contains more than 61,000 entries for 70 different elements, including latitude and longitude, depth, geographic region, jurisdiction, site description (geology and biology), types of hydrothermal activity, mineral deposit description, tectonic setting and bibliographic references;

(b) A method table containing information on methods used for analysing the different sample suites. For each sample an analytical method is indicated and referenced to a protocol in the table. The analytical method used for each element is listed together with the reported or inferred detection limit for that element by the specified method. A total of 110 different analytical protocols employed by about 23 different institutions that have published geochemical data on seafloor sulphides are listed. Yet of the different methods listed in the table, four techniques account for the majority of the data. These are atomic absorption spectrometry, optical emission spectrometry, X-ray fluorescence and instrumental neutron activation analysis;

(c) A submarine vents description set containing location, geological information and descriptions of 327 sites of seafloor hydrothermal activity and mineral deposits;

(d) A reference data set that includes citations of 540 literature references and other data sources that were used to construct the database.

14. The quality and quantity of data reported in the literature are highly variable, owing to different analytical methods, detection limits, sample size, and different standards for reporting of data. A major part of the compilation was to standardize the presentation of the data and to ensure that the compiled data are “sound”, including cross-checking references, making interlaboratory comparisons and developing a reporting protocol for different chemical data (e.g. detection limits for different elements and different analytical methods). The validation of the database was carried out by a qualified research scientist who inspected the data and decided what should be included and how it should be reported.

II. Status of implementation of the Central Data Repository

15. The Secretariat has developed the appropriate computer-based databases and progressively developed powerful and efficient interfaces to allow access via the Internet by representatives of member States, scientists, students and various other professionals.

A. Static tables

16. Initially, simple static tables were developed and posted on the Authority’s web site at www.isa.org.jm/data-rep/homepage.htm. These tables are still available and can be downloaded for analysis. They are as follows:

1. Polymetallic nodules

17. There are five data sets: sample data, cruise data, major elements, geochemical data and subsample comments. Each data set is divided into a number of tables, the main criteria being the ease of downloading and data manipulation. Over 80 tables have been created and are accessible via the Internet. These tables can be saved as HTML files or opened in a Microsoft Excel worksheet (the possibility also exists for downloading entire data sets for those with appropriate transmission capacity).

18. A unique identifier called “CDR sequence number” links all data tables and can be used for cross-data-set queries. The identifier is prefixed as follows:

"CDRNnnnnnn" for polymetallic nodules entries, for example "CDRN000001"

"CDRCnnnnnn" for cobalt-bearing ferromanganese crust entries

"CDRCRnnnnnn" for reduced data entries for the ferromanganese crusts, for example "CDRCR000020"

"CDRSnnnnn" for the polymetallic sulphides

19. Seven look-up tables provide description for the various codes used in the tables.

2. Cobalt-bearing ferromanganese crusts

20. Four data sets are provided: geochemical, major elements, reduced data and sample data. The geochemical (36 tables) and major element data sets (9 tables) are similar to those specified for the polymetallic nodules; they contain 3,533 records each. The sample data set (9 tables) contains ancillary data describing the sources and the characteristics of the primary data sets. Each is equally divided into a number of tables (here again, the entire data sets are available for download). The reduced data set (15 tables) contains a single entry for each location represented in the data set. The original data set includes multiple analysis for single sample collections. This data set averages together the data for all such replicate data, thus reducing the size of the original data set (1,225 records). It contains the same fields as the geochemical data set.

21. Two static maps showing the distribution of sample collection are available, one for the nodules and the other for the crusts.

3. Hydrothermal vent systems and sulphides database

22. The tables are based on the four data sets specified above. The geochemical data set has been split into subsets of 19 HTML static tables for the different geographic regions represented in the database. An interactive reference map is linked to these tables; by clicking on a specific region of the interactive maps, a link is created to a subset data table containing the geochemical data for deposits in that area. Most samples have data listings for at least 5 to 10 key elements (e.g., Cu, Fe, Zn, Pb, S, Au and Ag). However, data have also been compiled for each of the following elements: Fe, Cu, Zn, Pb, Au, Ag, Mn, As, Bi, Be, Cd, Co, Cr, Ga, Ge, Hg, In, Mo, Ni, Rb, Sb, Se, Sn, Sr, Sc, Te, Tl, U, V, W, Y, Zr, S, Si, Ba, Ca, C, Al, Mg, Ti, Na, K, P, Ir, Pd, Pt, Cl, F, B, Br, Hf, Li, Nb, Ta, Th, La, Pr, Nd, Sm, Eu, Gd, Tb, Dy, Ho, Er, Tm, Yb and Lu. In all cases, the methods used for each analysis are indicated in a "Method" column and described according to the key in the method table. Major elements are reported in weight per cent (wt %). Minor and trace elements are reported in parts per million (ppm) or parts per billion (ppb), as indicated. Non-metals are normally reported with the major elements as wt % oxides. All other elements are reported as elemental concentrations.

23. The other tables are based on the analysis method, description of vents systems and bibliographical references.

24. Based on past experience, this database is expected to increase in size by about 10 per cent per year, as new deposits are discovered and new data published. It will be updated accordingly.

B. Dynamic web interface

25. More recently, the Secretariat has acquired Oracle 8i, a web-enabled relational database management system. The system has been installed and configured, the database has been implemented and dynamic interfaces have been developed. The current technical environment for the CDR is as follows:

(a) Database server: Dell PowerEdge 2400, a Pentium III lower-end server;

(b) Map server: Dell PowerEdge 600Sc entry-level Pentium 4 server;

(c) Database management system: Oracle 8i, release 3;

(d) Application server: Oracle 9ias application server;

(e) Web interface: Oracle Portal, release 1;

(f) Mapping engine: MapInfo MapX 5.0;

(g) Mapping application server: MapInfo MapXtreme 3.0.

26. The Secretariat is connected to the Internet via a 256k leased line. The databases may be accessed via links from the Authority's web site (www.isa.org.jm) or directly at www.cdr.isa.org.jm. This offers a simplified interface to access not only the database but other information available at the web site. Specifically the CDR is accessed via the tab "Marine Resource Databases" and various sections are displayed. Three sections have been developed, respectively, for:

(a) Polymetallic nodules database;

(b) Cobalt-bearing ferromanganese crusts database;

(c) Seabed patents database.

The hydrothermal system and polymetallic sulphide databases are accessed via the tab "Sulphide Database/Seafloor Hydrotherm". In each section, there are links allowing each user to customize queries in the database and access specific data sets, such as cruise data, sample data, major elements data and all the geochemical elements. The user has the possibility to define any search criteria related to the geographic location (latitude, longitude, area name), the water depth, the thickness of crusts or the abundance of various geochemical elements. The following logical operators can be used: "=", ">", ">=", "<", "<=", "not null", "in", "not in", "null", "like", "!=" (different from). Boolean searches can be applied to any field or combination of fields. Indeed, this is in fact a powerful data analysis tool that allows a high level of flexibility. For example a user may wish to display within a defined area using either "Area Name" or "latitudes/longitudes" all locations where the iron content is >=25%, zinc>=0.1% and Ni>=0.4% with water depth <=1000m, etc. The result of such query may further be sorted according to any combination of up to six fields, for example iron content, latitude/longitude, water depth, thickness of crusts, etc.

27. Users may even specify the number of rows that a particular query will display. The default is set to 30, but depending on the capacity of his computer and Internet connection, the user may set it to hundreds or even thousands. Setting it to a higher number will take longer to display the records. It is important to understand that these are not static tables, but the result of queries from a database, a procedure which is much more complex than just accessing static pages. Here the tables are

built dynamically according to the criteria that the users define or according to the default criteria. Users have many possibilities for customization.

28. From any specific table, one can link to other tables. Development is under way to allow for cross-table queries. The use of this system is simple and intuitive, yet online help is available where appropriate and more development is under way to provide more of such help. A major phase is the ongoing development of dynamic graphical interfaces to provide interactive spatial presentation of the data. Such development will usefully complement and enhance the databases, as mapping is a powerful data presentation and analysis tool. This should complete the cycle in terms of functions and constitute, with the dynamic database, a full representative set illustrative of the entire project.

29. A background summary associated with each resource type is available online; it is in fact a full documentation of the data sets and provides an understanding of the overall analysis performed by the various expert consultants involved in the project. Users are encouraged to consult that documentation.

III. Seabed patents database and CD-ROM

30. Deep seabed mining is a formidable task. Developers of nodule mining technology need to address the basic question of how to pick up the nodules from the ocean floor and bring them to the surface. During the past 40 years, three basic design concepts for mining technology have been pursued: picking up nodules with a dredge-type collector and lifting them through a pipe; picking them up with a bucket-type collector and dragging the bucket up with a rope or cable; and picking up nodules with a dredge-type collector and allowing the collector to ascend by the force of its own buoyancy. In an effort to assist in the further development of technologies for exploration of polymetallic nodules in a rational manner, the Secretariat commissioned a survey of international patents to identify trends in the development of deep seabed mining technology covering the period from 1960 to 1998. The survey sought to define the state of the art in deep seabed mining as it relates to polymetallic nodule exploitation and to analyse the patent data with a view to identifying trends in seabed mining technology. Although not all technologies for deep seabed mining have been patented, the public record enables an examination of the evolution of the technology through time and the identification of key contributors in the field. The survey resulted in the identification of 352 patents from 12 patent systems. The majority of patents issued (85 per cent) were from the United States of America, Japan and the former USSR. The search focused on recovery technologies. Research and development activities commenced in the 1960s, peaked in 1983 with 34 patents issued and today continues at a much reduced pace. The Authority has compiled all that information in two CDs which allows easy access, searching with background information and statistical analysis. The CDs are readily available from the Secretariat. A summarized version of the seabed patents will be made available on the CDR.

IV. Library catalogue

31. The CDR also provides an interface to the Authority’s library catalogue. Over 900 books related to the law of the sea, marine mineral resources and other relevant

articles may be searched directly via the Internet, with criteria such as “Author”, “Language”, “Publication” and free text search on the title and subject. Other links to the main International Seabed Authority (ISA) page are being developed and will provide a unified interface, simplifying browsing to official documents, press releases, publications and the “What’s new” page.

V. Further development

32. Over the next two years, the Secretariat will pursue the development of the CDR in the following areas:

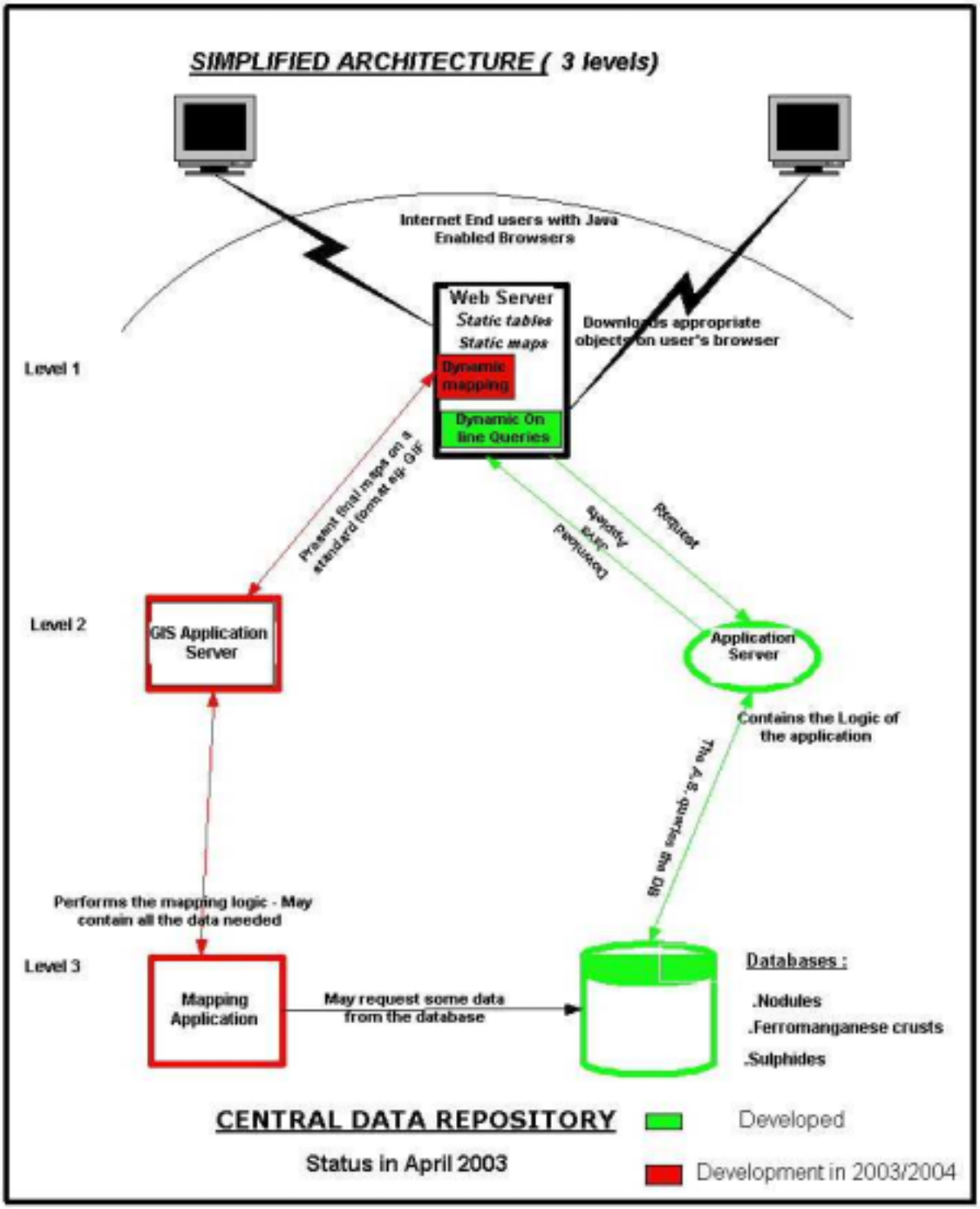
(a) Development and integration of graphical interfaces to provide visual data analysis tools via GIS over the Internet;

(b) Installation of a mirror site in the United States;

(c) Resumption of the collection of nodule data from other organizations/companies and their integration into the database structure;

(d) Development and integration of an environmental/biological database.

Annex


SIMPLIFIED ARCHITECTURE (3 levels)
Internet End users with Java Enabled Browsers
Web Server
Static tables
Static maps
Dynamic mapping
Dynamic On line Queries
Downloads appropriate objects on user's browser
Level 1
Present final maps on a standard format eg: GIF
Download Java Applets
Request
Level 2
GIS Application Server
Application Server
Contains the Logic of the application
The A.S. queries the DB
Performs the mapping logic - May contain all the data needed
Level 3
Mapping Application
May request some data from the database
Databases :
.Nodules
.Ferromanganese crusts
.Sulphides
CENTRAL DATA REPOSITORY
Status in April 2003
Developed
Development in 2003/2004