



Legal and Technical Commission

Distr.: General
24 June 2016

Original: English

Twenty-second session
Kingston, Jamaica
11-22 July 2016

Data management strategy of the International Seabed Authority

I. Introduction

1. The Legal and Technical Commission, in its report to the Council in 2015, requested the secretariat to provide a draft data management strategy and the financial implications of its implementation (ISBA/21/C/16, para. 36). In response to that request, the present document presents a summary of the preliminary review of the current data management arrangement by the secretariat that was prepared by a working group of the Commission and, subsequently, the data management strategy following the recommendations of an external expert consultant, INSPIRE Environmental, United States of America.

II. Legal and Technical Commission preliminary review of current data management arrangements at the secretariat

2. The Legal and Technical Commission formed a working group on data management strategy during its February 2016 session. The working group conducted a preliminary review of the current data and information management arrangements and of the existing hardware and infrastructure supporting that arrangement. The working group noted, regarding the secretariat, that:

(a) In the current management system, data and information are stored in four places, namely:

- (i) A secure vault for the initial registering and physical storage;
- (ii) A server that is divided between the public International Seabed Authority website and the Legal and Technical Commission secure website;
- (iii) A stand-alone computer with a geographic information system;
- (iv) A stand-alone computer with several software packages for handling biological/environmental data;



(b) There is no direct link between the four units; all uploading to the server is performed manually. Data are imported into the stand-alone units not as a routine, but according to day to day needs. The transfer of data from the biological/environmental stand-alone site to the geographic information system is performed manually;

(c) The basic archiving function, that is, bookkeeping, searching and retrieving of data and information, is also performed manually based on the content of the physical storage (vault). A selection of the information may be retrieved from the websites, but their content relies on the subjective choices of the individual webmasters and does not necessarily reflect the needs of the users. No one person seems to be designated to be in charge of the overall data management.

3. The working group concluded with the following observations:

(a) The Authority must implement routines and technical means to take care of all data and information in a structured manner consistent with all the needs of the Authority's functions. This includes the basic archiving needs and all the relevant uses of the data and information;

(b) A staff member in the position of data manager is needed to keep such a system working;

(c) More than one geographic information system officer is needed to prevent the system from becoming vulnerable;

(d) The staff computers should be upgraded according to their use and function.

4. The working group also stated that the setting up of the strategy comprises priority steps and progressive implementation aspects, in particular:

(a) Review of the current arrangements for the use of data by the Authority in the monitoring and evaluation of contractors' activities (environmental, geological and technological integration);

(b) Definition of a (an Authority) model of data management;

(c) Determination of the type of technology and related infrastructure to be used to implement the plan;

(d) Definition of a timeline showing the sequences of the implementation plan;

(e) Related financial evaluation;

(f) Setting up of the Authority's ameliorated system of data and information tool;

(g) Elaboration of policies and procedures for accessibility;

(h) Definition of the level of accessibility to the information by other stakeholders (external to the Authority).

5. The working group noted that the internal uses of the Authority (steps 4 (a)-(f) above) would have first priority and that the external uses (steps 4 (g) and (h)) would be considered subsequently.

III. Outcome of the data management consultancy

6. The consultant, INSPIRE Environmental, was engaged in April 2016 and mandated to recommend a data management strategy for the handling and the use of confidential and non-confidential data provided by exploration contractors to the Authority.

7. The consultant's scope, which was based on the aforementioned items identified by the working group in paragraph 4 above, was to:

(a) Conduct a review of the current status and arrangements for the use of data by the Authority in the monitoring and evaluation of contractor data (environmental, geological/exploration and technological integration);

(b) Define a suitable model of data management for the Authority. The primary users will be the secretariat and the Legal and Technical Commission; the secondary audience will be the interface with contractors and the general public;

(c) Determine the type of technology and related infrastructure required to support the suggested data management model;

(d) Define a timeline showing the sequences of the implementation plan;

(e) Provide financial implications of the data management model implementation (including contractor, human resources, initiation and annual operating costs).

8. INSPIRE Environmental conducted an on-site review of current practices from 22 to 26 April 2016 and provided recommendations for the data management system, including the type of technology and related infrastructure required to support it.

9. The data management system was assessed through on-site interviews with data managers and data users. The goal of the interviews was to gain an understanding of the current system, assess shortcomings in its structure and discover the needs for data management to support the recommendations, options and costs detailed in the consultant's report.

10. On the last day of the on-site visit, a meeting was held between INSPIRE Environmental and the Authority/Office of Resources and Environmental Monitoring staff. An outline of the consultant's report was presented. The policy for handling confidential data was discussed and an overview of the policy was incorporated into the preliminary data management strategy report. The final report was submitted on 10 May 2016.

IV. Recommendations

A. Data management plan: policies and procedures

11. In addition to the existing data management policies, the following should be considered in the data management plan:

(a) A time limit on confidentiality should be re-instigated for a period of time to be agreed upon when the new data management model is in place;

(b) The model must be able to answer basic questions in support of the Authority's mission, including estimating mineable areas, as well as evaluating environmental impacts of resource exploitation;

(c) The model must enable users to easily summarize the volume and summary content of data collected by the Authority and incorporated into the database (e.g., metadata);

(d) The model must have a clear, published and adhered to set of data standards.

B. Workflow

12. The data management plan will be informed by the ideal process for data flow at the Authority. There will be two main sources of data for the period of the transition, historical data and new data:

(a) **Historical data.** Historical data will be migrated from existing digital sources (digitized from pdf reports, as well as Oracle, Excel and other digital formats). This activity includes both confidential and non-confidential data. A procedure will be developed during the database development phase to load the data into the main database;

(b) **New data.** The vision for incorporation of new data into the database has the following features:

(i) A metadata form that must be filled out (required fields) by the contractor when submitting data. The content of the form will be designed to meet the criteria of the metadata catalogue of the database. The form will be designed as part of the task for creating the data management plan;

(ii) Standard Excel files will be submitted and checked for content and formats. The files will be modified as the database structure is generated;

(iii) Other digital files can be submitted in a variety of formats, as defined in the report of the Legal and Technical Commission on recommendations for the guidance of contractors on the content, format and structure of annual reports ([ISBA/21/LTC/15](#)).

13. The data will be managed by the Authority's data manager to ensure that:

(a) The confidential status of the data is maintained;

(b) The content of data submitted complies with the requirements;

(c) The data are correctly catalogued relative to the metadata;

(d) The database is regularly updated, as required, and kept current.

14. A process to load the data into the master database will be created during the database development phase. There may be additional new data directly incorporated into the master database from other sources.

15. Users, such as the Legal and Technical Commission, that need access to the contractor reports will be able to access them via a virtual private network or other secure mechanism.

16. The general public will be able to access environmental, bibliographic and other non-confidential data from the Authority website via the Internet.

C. Database model

17. The database will enforce valid values for standards generated for fields including chemical names and units, species names and DNA taxonomy. This will enable accurate and complete querying.

18. Established geographical standards will be documented and enforced.

19. The database model will also include the storage and retrieval framework for non-structured data (e.g., photos, videos, scanned graphics), which will be geo-referenced (if applicable).

D. Database and mapping software

20. Considering the geographical focus of the project, a spatial database (or geodatabase) will be selected as the database engine. The spatial database will be optimized to store and query data that represent objects defined in a geometric space.

21. There are two primary options for database software: proprietary and open source.

E. Infrastructure and hardware to support the data management strategy

22. **Existing infrastructure and hardware.** The current technical infrastructure in place at the Authority is a collection of older physical servers, aged switching and previous-generation server operating systems. While there are fundamental best practices at play, the results of the review indicated that the overall environment has reached the end of its life cycle and is in need of a replacement. Current generation concepts, such as virtualization for added management efficiency; flexibility, security, reliance and performance, should be implemented.

23. The recommendation is to use virtual servers wherever possible. Not only does a virtual environment provide far more flexibility than a physical environment, but it is also generally more cost-effective. It allows an organization to have high availability of its workloads, easy scaling and growth; it allows for single pane of management; and it allows for granular and complete backups to occur quickly, with quick recovery times in the event of issues. An enterprise-grade virtual environment is recommended.

F. Project phases and timeline

24. The data management strategy will be implemented over a period of 19 months, starting in January 2017. It will be organized under the nine tasks described below:

(a) **Project initiation.** A kick-off meeting on-site should be scheduled with the consultant in January 2017. The goals of the meeting are to review the scope, set priorities, set up communication lines and review the schedule for the project. Timeline: two working days in January 2017;

(b) **Data management plan.** A formal data management plan will be generated to define the policies and procedures for conducting the project. It will include diagrams of the data structure; plans for updating the templates; designing the user interface for access to the data; formalizing the method for migration of archival data to the database; formalizing standards and valid values; updating the data flow based on feedback from the Authority; and a description of the roles and responsibilities for data management staff. A draft plan will be ready for review and comment by the Legal and Technical Commission by the end of February 2017. The plan will then be finalized after taking into account the Commission's contributions. Timeline: three months (January-March 2017);

(c) **Information technology design, acquisition and deployment.** This task involves building the virtual environment hardware (storage, switching, servers) and includes time for the contractor information technology architect. Timeline: three months (January-March 2017);

(d) **Information technology installation and training.** This activity involves configuring the virtual environment, setting up procedures and training on-site staff. Timeline: two weeks (April 2017);

(e) **Database and user interface development.** This task involves designing and developing the following:

(i) The geodatabase structure;

(ii) A method for storage and retrieval of both structured and unstructured data;

(iii) A web-based user interface for the data via both map view and searchable catalogue interface;

(iv) A method for classifying data as confidential or non-confidential and being able to update this status;

(v) An update of the data flow for sharing the data with the Legal and Technical Commission as well as the public. Timeline: April-August 2017;

(f) **Data migration.** Historical data will be migrated into the database using a variety of methods depending on the data source. Timeline: April 2017-May 2018;

(g) **Database implementation, Testing and training.** This task includes time to load software on the information technology infrastructure installed at the site; move the database to the appropriate servers; set up procedures and train on-site staff for data loading, validation and website and related data maintenance procedures. Timeline: December 2017-February 2018;

(h) **Documentation.** This task concerns documentation on both information technology and database maintenance and involves information technology documentation to be given after deployment and database documentation to be given at the end of the project. Timeline: September 2017-February 2018;

(i) **First-year implementation support.** The aim of this task is to remotely support the information technology infrastructure following installation while the database is being developed (first year) and to provide support for issues that arise as the database, map server and updated website are deployed. Timeline: July 2017-June 2018.

G. Preliminary estimated costs

25. The implementation costs for the recommended data management strategy for the biennial budget period of 2017-2018 are estimated at \$675,528 (see annex). They include the hardware acquisition, infrastructure deployment, software development and first-year support and operational costs.

26. After 2018, the yearly operational costs are estimated at \$35,872.

H. Additional staffing requirement

27. Implementation of the data management strategy requires two new posts: a Data Manager (Professional) and a Data Entry Officer (General Service). The new posts would be budgeted under the human resources established posts item of the Authority.

Annex

Proposed cost estimate for implementation of the data management plan and infrastructure upgrade

(United States dollars)

<i>Task No.</i>	<i>Task name</i>	<i>Year</i>	<i>Consultant labour</i>	<i>Annual maintenance and support</i>	<i>Hardware, software and information technology infrastructure</i>	<i>Consultant travel</i>	<i>Total</i>
			Charged to new annual	Charged to consultants budget line for programme 2.4	Charged to annual maintenance and support budget line for programme 2.4	Charged to information technology budget line	Charged to travel budget line for programme 2.4
1.	Project initiation	2017	21 312	–	–	12 000	33 312
2.	Data management plan	2017	23 772	–	–		23 772
3.	Information technology design	2017	51 952	–	–	–	51 952
4.	Information technology installation, testing and training (hardware option 3)	2017	22 720	–	250 000	4 000	276 720
5.	Database and user interface development	2017	40 632	–	50 500	0	91 132
6.	Data migration	2017: 50 per cent; 2018: 50 per cent	20 352	–	5 000	8 000	33 352
7.	Database implementation and testing	2017	23 312	–	–	4 000	27 312
8.	Documentation	2017	22 072	–	–	–	22 072
9.	First-year implementation and support	2017	80 032	–	–	–	80 032
Subtotal			306 156	–	305 500	28 000	639 656
10.	Annual support, starting in 2018	2018	–	35 872	–	–	35 872
Total, 2017-2018			306 156	35 872	305 500	28 000	675 528

Assumptions and notes

1. Human resources costs are assumed to be included in the budget line for established posts.
2. The consultants will not charge overhead, as this will be borne by the Authority as part of established posts, including human resources, finance, and procurement.
3. Source is the INSPIRE Environmental consultants report of May 2016.